Molecular modeling of bacterial polysaccharide antigens to inform future vaccine development

Author: **Jason Hlozek**

Supervisors: Neil Ravenscroft, Michelle Kuttel

Thesis Presented for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Chemistry

UNIVERSITY OF CAPE TOWN

October 2020

Declaration

Molecular modeling of bacterial polysaccharide antigens to inform future vaccine development

I, Jason Peter James Hlozek, declare the following:

1. That the above-titled thesis is my own work, both in conception and execution, apart from the normal guidance of my supervisor(s);
2. That in cases where others' work has been cited, this has been acknowledged and referenced;
3. That no part of this thesis has been submitted in the past, or is being, or is to be submitted for another degree at this or at any other University;
4. That I grant the University of Cape Town free license to reproduce this thesis in whole or in part, for the purpose of research.

I hereby present this report in fulfilment of the requirements for the degree of Doctor of Philosophy in the Department of Chemistry at the University of Cape Town.

Signed by candidate

_____

Doctoral Candidate's Signature

October 2020

I confirm that I have been granted permission by the University of Cape Town's Doctoral Degrees Board to include the following publication(s) in my PhD thesis, and where co-authorships are involved, my co-authors have agreed that I may include the publication(s):

a. Hlozek, J.; Kuttel, M. M.; Ravenscroft, N., Conformations of *Neisseria meningitidis* serogroup A and X polysaccharides: The effects of chain length and O-acetylation. *Carbohydr. Res.* **2018**, *465*, 44-51.

b. Hlozek, J.; Ravenscroft, N.; Kuttel, M. M., Modeling the conformations of *Neisseria meningitidis* serogroup a CPS and a carba-analogue: Implications for vaccine development. *Carbohydr. Res.* **2019**, *486*, 107838.

c. Hlozek, J.; Ravenscroft, N.; Kuttel, M. M., Effects of Glucosylation and O-Acetylation on the Conformation of *Shigella flexneri* Serogroup 2 O-Antigen Vaccine Targets. *J. Phys. Chem. B* **2020**, *124* (14), 2806-2814.

d. Hlozek, J.; Owen, S.; Ravenscroft, N.; Kuttel, M. M., Molecular Modeling of the *Shigella flexneri* Serogroup 3 and 5 O-Antigens and Conformational Relationships for a Vaccine Containing Serotypes 2a and 3a. (Submitted, Under review).

Signature: Signed by candidate Date: 16 September 2020

Student Name: Jason Hlozek Student Number: HLZJAS001

Table of Contents

Abstract

Polysaccharide conjugate vaccines have been pivotal in reducing the prevalence and severity of bacterial infectious diseases worldwide, preventing countless deaths. The effectiveness of a vaccine can be extended if the selected vaccine strains in a multivalent vaccine cross-protect against non-vaccine strains. Detailed knowledge of antigen structure and conformation is required for vaccine components to be rationally selected. However, experimental methods may not be able to ascertain the conformations of polysaccharide chains. To address this, molecular dynamics simulations can provide key theoretical insights on molecular conformation to rationalize cross-protection data and inform vaccine development.

In this work, we use molecular dynamics to investigate the conformations of glycan antigens of *Neisseria meningitidis* and *Shigella flexneri* bacteria - causative agents of meningitis and diarrheal disease. For *N. meningitidis*, our modeling indicates why vaccination with serogroup A has not cross-protected against serogroup X infection, justifying the inclusion of serogroup X in future multivalent meningococcal vaccines. We also find that a chemically-stable carba-analogue of serogroup A has significant conformational differences to the native serogroup A chain, which does not support its use as a suitable serogroup A vaccine replacement. Our simulations of *S. flexneri* glycan antigens (serogroups Y, 2, 3, and 5) identify heuristics for the effects of substitution on backbone conformation and supports a proposed vaccine containing serotypes 2a (with O-acetylation) and 3a that will provide broad cross-protection. These findings can guide the rational selection of vaccine components to result in next-generation vaccines with greater cost-effectiveness and improved disease coverage.

Abbreviations

CPS	Capsular Polysaccharide
LMICs	Low- and Middle-Income Countries
LPS	Lipopolysaccharide
MD	Molecular Dynamics
MenA	*Neisseria meningitidis* serogroup A
MenX	*Neisseria meningitidis* serogroup X
NMR	Nuclear Magnetic Resonance
O-Ag	O-antigen
PMF	Potential of Mean Force
RU	Repeating Unit

Acknowledgments

A massive thank you to my supervisors for many years of mentorship, support and amazing opportunities – I am very grateful for all your input during this PhD degree. Assoc. Prof. Michelle Kuttel, thank you for motivating me to achieve more than I thought I was capable of and for your thorough work reviews that always ensured research outputs of the highest quality. Assoc. Prof. Neil Ravenscroft, thank you for your constant support, guidance, enthusiasm, and ever-present humor. A further thank you to Prof. Graham Jackson for stimulating and insightful discussion.

Thank you to my mother, Michelle Hlozek, and late grandfather, Peter Best, who were always my biggest supporters and the proudest of my achievements. To my partner, Sarah Milborrow, and many friends: thank you so much for your support and for going far above and beyond in trying times - I will forever be grateful.

Thank you to the NRF, UCT, Department of Chemistry, and my supervisors for funding this work as well as various conference travels. Thank you to the administration staff of the Chemistry Department for always being willing to help and to go the extra mile. Finally, thank you to the ICTS-HPC team for your dedication to maintaining the HPC cluster on which this work was computed, and for your willingness to help whenever needed.

List of publications

This dissertation is a thesis with the inclusion of publication. Table 1 lists the four publications that form a part of this dissertation. Table 2 includes a further co-authored publication that is not included in this thesis.

Table 1: List of publications included in this dissertation

Title	Status	Journal	Authors	Contributions	Ref
Conformations of *Neisseria meningitidis* serogroup A and X polysaccharides: The effects of chain length and O-acetylation.	Published	Carbohydrate Research	Hlozek, J., Kuttel, M. and Ravenscroft, N.	Hlozek provided the simulation data sets for the strains in the study: MenA/ MenA3Ac/MenX. Hlozek performed the analysis of the simulations and NMR data. Further, Hlozek conceived and drafted the paper, including the figures. Ravenscroft provided the NMR spectrum for MenX. Kuttel and Ravenscroft reviewed and edited the article.	1
Modeling the conformations of *Neisseria meningitidis* serogroup A CPS and a carba-analogue: Implications for vaccine development.	Published	Carbohydrate Research	Hlozek, J., Ravenscroft, N. and Kuttel, M.	Hlozek provided the simulation data sets for MenA (extended simulation) and Carba-MenA. Hlozek performed the data analysis for both simulations. Further, Hlozek conceived and drafted the article, including the figures. Kuttel and Ravenscroft reviewed and edited the article.	2

Table 1 (continued):

Effects of Glucosylation and O-Acetylation on the Conformation of *Shigella flexneri* Serogroup 2 O-Antigen Vaccine Targets.	Published	Journal of Physical Chemistry B	Hlozek, J., Ravenscroft, N. and Kuttel, M.	Hlozek provided the simulation data sets for the four *Shigella* serotypes in the study: 2a/2a-3Ac/2b/Y. Hlozek conceived and drafted the article. Hlozek performed the data analysis for all simulations. Further, Hlozek conceived and drafted the article, including the figures. Kuttel and Ravenscroft reviewed and edited the article.	3
Molecular Modeling of the *Shigella flexneri* Serogroup 3 and 5 O-Antigens and Conformational Relationships for a Vaccine Containing Serotypes 2a and 3a.	Under Peer Review	Manuscript included here	Hlozek, J., Owen, S., Ravenscroft, N. and Kuttel, M.	Hlozek conceived and drafted the article, including the figures. Owen provided insights and initial data sets for the serogroup 3 component with Hlozek providing new data sets for serogroups 3 and 5 (3a/3b/5a/5b). Hlozek performed the data analysis for all simulations. Kuttel and Ravenscroft reviewed and edited the article.	N/A

Table 2: Co-authored publication not included in this dissertation

Title	Reason not included	Journal	Authors	Contributions	Ref
Genetic and structural elucidation of capsular polysaccharides from *Streptococcus pneumoniae* serotype 23A and 23B, and comparison to serotype 23F.	Hlozek is the 3rd author. Although it considers a similar vaccine context, this article arose from a large collaboration and does not consider molecular simulations.	Carbohydrate Research	Ravenscroft, N., Omar, A., Hlozek, J., Edmonds-Smith, C., Follador, R., Serventi, F., Lipowseky, G., Kuttel, M., Cescutti, C., Faridmoayer, A.	Hlozek provided NMR assignment and analysis for the serotype 23B component of the article.	4

Conference presentations

Parts of the work included in this dissertation have been presented at the following conferences:

- Capsular polysaccharide conformations in meningococcal serogroups A and X (Conference Poster Submission), Hlozek, J., Ravenscroft, N., Kuttel, M.M. (2017). EuroCarb2017 Conference, 2-6 July 2017, Barcelona, Spain.
- O-Antigen polysaccharide conformations of *Shigella flexneri* serogroups 2 and 5 (Conference Poster Submission), Hlozek, J., Ravenscroft, N., Kuttel, M.M. (2018). ICS2018 Conference, 14-19 July 2018, Lisbon, Portugal.
- *Shigella flexneri* Lipopolysaccharide Conformations: Effect of glucosylation and O-acetylation in serogroup 2 (Oral Communication - Abstract 224), Hlozek, J. Glyco25, XXV International Symposium on Glycoconjugates, (2019). Glycoconjugate Journal,36, 327-328.

Chapter 1

Introduction

1.1 The effectiveness of vaccines in controlling disease

Vaccination is a fundamental medical practice that reduces the morbidity and mortality rates of infectious diseases.[5, 6] Vaccines prevent approximately 2.5 million deaths per year[7] and typically reduce disease severity if subsequent infection occurs.[8] In previous decades, significant progress has been made in the number of diseases prevented by licensed vaccines as well as in the development of new vaccination technologies.[9] However, vaccines do not always reach the most affected populations – approximately 20% of deaths in children under five years of age could be prevented by currently available vaccines.[10-12]

These vaccine preventable diseases disproportionately affect low- and middle-income countries (LMICs) who often lack adequate health care infrastructure or the socioeconomic means to purchase vaccines.[13] The greatest burden falls on the African continent where vaccination rates for common infectious diseases are typically below the global average[14] and with a corresponding child mortality rate from vaccine preventable diseases accounting for as much as 50% of the global total.[15] These diseases are becoming notably more serious with increasing development of antibiotic resistance, which renders many conventional treatments ineffective.[16, 17] Therefore, the development of inexpensive and effective vaccines that prevent disease are key to reducing the extensive disease burden on Africa. In particular, the bacterial pathogens

that contribute to meningococcal and diarrheal diseases have long been major contributors to the severity of the African disease burden and are addressed specifically in this dissertation.

Vaccines protect an individual that is susceptible to a particular disease by introducing a weakened form of that pathogen or a smaller constituent thereof to elicit a protective immune response.[12] This trains the immune system to recognize the pathogen and it is subsequently primed to defend against natural infection if the pathogen is encountered in future. For infectious bacteria, vaccines commonly target the glycans expressed on the bacterial cell surface. These polysaccharides have tremendous structural variety given the large number of possible monosaccharide building blocks as well as the multiple hydroxyl groups through which the monosaccharides can be linked together.[18] The unique polysaccharide chains are typically used as the defining antigen for classifying bacterial strains into serotypes and serogroups. These polysaccharide antigens can be obtained through purification from bacterial cultures, chemically synthesized, or made biosynthetically by recombinant expression.[19] Modern vaccines further enhance the immunogenicity of polysaccharide components through conjugation to an antigenic protein carrier.[20]

To date, significant progress has been made towards the development and licensure of bacterial polysaccharide-based vaccines.[19, 21] Diseases are typically caused by several strains that circulate simultaneously in a given region, such as *Neisseria meningitidis*, *Shigella flexneri*, *Streptococcus pneumoniae*, *Klebsiella pneumoniae*, etc. Therefore, vaccines often need to consist of several serogroups/serotypes (vaccine valency) to provide broad protection. A key challenge in vaccine development is to design a

vaccine that provides maximal coverage against prevalent bacterial strains with minimal vaccine valency, reducing vaccine cost. The valency of a vaccine can be reduced if the selected vaccine serotype(s) can confer protection against closely related non-vaccine serotypes.

Serotype cross-protection typically occurs between chemically similar strains within the same serogroup; however, similarity in primary structure alone is not a reliable predictor for cross-protection.[22] For example, the first seven-valent vaccine against *Streptococcus pneumoniae* (PCV7) included serotype 19F and cross-protection was expected against the related serotype 19A. Both serotypes express a similar trisaccharide repeating unit (RU) with a difference in a single linkage position, yet 19A infection increased post-PCV7 vaccination[23] and only limited cross-protection was observed from PCV7 in clinical study.[24] Computational modeling could rationalize these findings with the discovery of conformational differences between the serotypes: 19F had a higher prevalence of extended conformations, while 19A instead adopted more compact structures with hairpin bends.[25] Therefore, antibodies raised against the 19F chain are unlikely to recognize and bind to the conformations adopted by 19A with the same avidity, resulting in diminished cross-protection against 19A and explaining the increase in 19A infection after PCV7 vaccination.

A second example is that of *Neisseria meningitidis* serogroups Y and W. Both serogroups have hexose-sialic acid RUs where the hexose residues are C-4 epimers (glucose for Y compared to galactose for W). Despite no structural change at a linkage position, clinical study reports unusual asymmetric cross-protection: vaccination with serogroup Y elicited a cross-protective antibody response against serogroup W in 71%

of volunteers, while vaccination with serogroup W only elicited cross-protective antibodies against serogroup Y in 30% of volunteers.[26] Computational modeling of the polysaccharides chains revealed a dramatic difference in conformations: serogroup Y had a single dominant conformation while serogroup W instead adopted a range of conformational families, which included the dominant conformation of serogroup Y. These findings could rationalize the clinical data - the majority of antibodies raised against Y could also cross-react with a conformation adopted by W, however, only a small proportion of antibodies raised against the array of conformations expressed by W would recognize the single conformation of Y.[27] Therefore, as even small structural changes can lead to significant differences in polysaccharide conformation and flexibility, a central assumption for cross-reactivity is that both chemical and conformational similarity are required for effective cross-protection to occur.

A detailed understanding of antigen structure and conformation enables a rational approach towards the selection of suitable vaccine components in the vaccine design process. For glycans, the most common experimental methods for obtaining structural data are X-ray crystallography and Nuclear Magnetic Resonance (NMR) spectroscopy. However, although powerful, these methods are limited in the scope of information they provide. X-ray crystallography relies on the ability of the polysaccharide to crystallize (or co-crystallize with a bound antibody) and represents the glycan conformation in the solid phase, however, this may not be representative of the potential range of conformations adopted in solution. NMR is a powerful technique for determining molecular structure, yet, the resulting spectra represent the dynamic

average of polysaccharide conformations in solution and may be ambiguous for determining the conformation of flexible glycan chains.

Computational modeling is increasingly employed to gain theoretical insight into glycan conformation, which may otherwise be inaccessible to experiment. As we transition from a history of empirically designed vaccines toward an age of rational vaccine design,[28, 29] molecular modeling in combination with NMR data can be a powerful methodology to rationalize empirical cross-protection findings and guide future glycoconjugate vaccine design.

1.2 Molecular dynamics theory

Molecular simulations aim to reproduce the time dependent behavior of microscopic molecular systems in order to explain and predict physical phenomena. A molecular system can be fully described by solving the quantum mechanical wave functions of the system components; however, this is often computationally infeasible to compute even for small solvated systems. Instead, classical physics descriptions can be employed, which are much faster to compute and can sufficiently approximate the physical properties of much larger systems.[30]

This work makes use of molecular dynamics (MD) simulations to provide detailed atomistic information of glycan antigen conformations. MD explicitly models each atom in the system as a charged point-mass whose interactions with other atoms are described by classical statistical mechanics. In particular, atom movement is governed by Newton's second law of motion (Equation 1.1), from which the Newtonian equations of motion are derived.

$$F = m_i a_i \qquad (1.1)$$

A force, F, acting on an atom is related to its potential energy, U, by equation 1.2. Therefore, equation 1.1 and 1.2 can be combined to allow an atom's movement to be computed from its potential energy (Equation 1.3).

$$F = -\frac{\partial}{\partial r_i} U \qquad (1.2)$$

$$m_i a_i = -\frac{\partial}{\partial r_i} U \qquad (1.3)$$

The potential energy for each atom is calculated as a sum of energy terms corresponding to its bonded interactions and non-bonded electrostatic contributions (Equation 1.4). The specific combination of equations and constants used to represent this potential energy is termed a force field. A force field aims to produce MD simulations that closely match the behavior of a target set of data (typically quantum mechanical or empirical data). The publications presented in this thesis make use of the CHARMM 36 all-atom parameter set,[31, 32] parameterized against quantum mechanical calculations, with force field equations defined within the NAMD software.[33]

$$U_{total} = U_{bond} + U_{angle} + U_{dihedral} + U_{Urey-Bradley} + U_{vdW} + U_{Coulomb} \qquad (1.4)$$

In general, a MD simulation produces a time-dependent trajectory by iterating through the cycle calculations in the cycle indicated in Fig. 1.1. First, the total potential energy is calculated for each atom in the system (according to Equation 1.4), which is then related to a net force acting on that atom. This net force updates the corresponding velocity and acceleration for each atom. Each atom is then moved according to

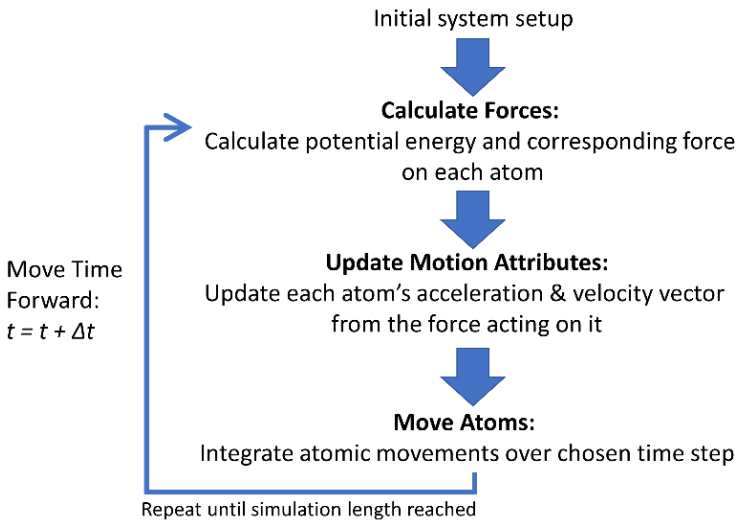


Figure 1.1: Iterative calculation to compute molecular dynamics simulations.

Newton's equations of motion over a pre-defined time step and this process is repeated until the desired simulation length is reached.

A simulation can approximate the time-dependent motion of each atom in the system, provided the time step is chosen to be sufficiently less than the smallest molecular motions. The smallest molecular movements are bond length vibrations that occur on the order of ~10 femtoseconds (10×10^{-15} s)[34] and, thus, a timestep on the order of 1 femtosecond is typically chosen for MD simulations. Therefore, a microsecond-length simulation requires 1 billion (1×10^{9}) iterations of this algorithm and also scales exponentially with system size due to the n-body nature of electrostatic potential calculations. Understandably, this is time intensive to compute - even a supercomputer with algorithmic optimizations can take months to complete a single simulation.

A key question to then consider is: at what point is the simulation length deemed long enough in order to draw meaningful conclusions from the data? Grossfield et. al. acknowledges the difficulty in addressing this question as all simulations have inherent

statistical uncertainty and, therefore, can never be considered absolutely converged.[35] Compounding the issue, no amount of analysis can reveal whether an important conformation remains unvisited in a simulation unless the conformation is already known by other means. However, several visual and qualitative techniques can be used to assess whether a simulation has *not* run long enough and can build confidence in the final data set.

In this work, we inspected plots of various metrics as time series (root-mean-square deviation, end-to-end distance and torsion angles) to ensure that simulations were not terminated whilst exploring novel conformations and that dominant conformations had been sampled multiple times. Further, potential of mean force (PMF) plots for each disaccharide linkage had been calculated prior to each simulation and the low-energy regions could be compared against the trajectory values. Significant differences in conformational occupancies could be used to indicate the extent of overall sampling or otherwise highlight conformational aspects requiring further investigation. Furthermore, the methodology employed here of comparative study between related glycans under identical modeling parameters provides insight into relative conformational differences and, therefore, partially mitigates the impact of absolute simulation length.

At the time of publication of the initial two journal articles in this dissertation, the *N. meningitidis* simulations studied were the longest computationally feasible data sets to compute and employed predominantly qualitative assessment of the conformational sampling. For the *S. flexneri* serogroups in the final two publications, additional statistical quantitative methods were employed in the form of block averaging analysis

to build further confidence in the conformational sampling. Provided that simulations of a reasonable length can be feasibly calculated, a comparative study of the dynamic motion of modeled polysaccharides can confidently rationalize and potentially predict cross-protection phenomena based on conformational arguments.

1.3 Computational methodology

This work employs an incremental systematic approach toward the computational study of closely related glycan antigens. First, we calculate the preferred glycosidic linkage orientations of each disaccharide fragment in the glycan antigen. These are determined by computing the potential of mean force of rotation about the linkage torsion angles (ϕ and ψ) with a metadynamics routine.[36] The resulting Ramachandran-like contour plot identifies favorable low-energy linkage conformations.

Next, short glycan chains (6-12 sugar residues) are built with linkage conformations corresponding to the energy minima identified from the PMF calculations. MD simulations are then performed on the solvated short-chains to identify favored backbone conformations in solution.

Finally, longer polysaccharide chains are built from the most favored linkage conformations in the short-chain MD simulations. MD simulations for the long polysaccharide chains are then run for a long simulation length (typically a microsecond) to obtain a data set or “production run” that approximates the native backbone conformations. The final data sets can be probed for convergence and extended if needed. Broad insights are gleaned from the glycan simulations by plotting the time trajectories of the end-to-end distances as well as scatter plots of the linkage

torsion angles. Clustering analysis is employed to determine and compare the dominant chain conformations. The research here is conducted in the format of comparative study between antigens that are treated under identical modeling conditions. This reveals the relative differences in conformational behavior between polysaccharide antigens as induced by small structural alterations in their respective RUs.

1.4 Aims

This dissertation includes four journal articles that aid the development of optimized vaccines for diseases that are particularly prevalent in LMICs on the African continent and further detailed literature is included within each publication. The first two papers investigate potential antigens for use in next-generation multivalent vaccines against *Neisseria meningitidis* (meningococcal) infection. The next two papers focus on computational modeling of candidate vaccine serotypes to prevent diarrheal disease caused by *Shigella flexneri* infection and explore the potential cross-protection of a proposed multivalent vaccine.

The dissertation therefore takes the form of a set of case studies that rationalize vaccination cross-protection data if available, or otherwise provides predictive insight into glycan antigens to guide vaccine design.

Therefore, this work broadly aims to contribute a part of the answer as to how molecular dynamics simulations of glycan antigens can be leveraged to provide key theoretical insights that contribute towards the rational basis of vaccine antigen selection.

1.5 Dissertation structure

This is a thesis that includes publications, as listed in Table 1.1. The first two of four results chapters, Chapter 2 and 3, consider the modeling of *N. meningitidis* polysaccharide antigens to guide the development of next-generation meningococcal vaccines.

Chapter 2 investigates whether a meningococcal serogroup A vaccine may confer cross-protection against the related serogroup X and, therefore, whether a serogroup X component is recommended for inclusion into current multivalent vaccines in development. It continues by determining the effect of the O-acetylation moiety on the conformation of the serogroup A backbone.

Chapter 3 analyzes the conformational differences between the native meningococcal serogroup A polysaccharide and a carba-analogue polysaccharide that is proposed as a chemically-stable alternative to current serogroup A vaccines.

The second section of this dissertation examines the O-antigens (O-Ags) of *S. flexneri* strains, to inform current development efforts toward the first licensed *Shigella* vaccine.

Chapter 4 investigates the conformations of serogroup 2, the dominant cause of shigellosis infection, and the unsubstituted serotype Y backbone.

Chapter 5 continues by exploring serogroups 3 and 5 to glean potential cross-protection relationships and to provide support for the composition of a proposed multivalent vaccine containing serotypes 2a and 3a.

Chapter 6 concludes the dissertation by providing an overview of the preceding chapters as well as discussing further research questions that arise from this work.

Chapter 2

Modeling the CPS of *N. meningitidis* serogroup A and X

2.1 Foreword

The following article "*Conformations of* Neisseria meningitidis *serogroup A and X polysaccharides: the effects of chain length and O-acetylation"* investigates two strains of the causative agent of frequent meningitis epidemics in Africa – *Neisseria meningitidis*. Past African meningococcal epidemics were predominantly due to serogroup A, with vaccination playing a pivotal role in controlling the disease burden. The affordable serogroup A vaccine developed for Africa (MenAfriVac) was deployed in mass vaccination campaigns from 2010, which subsequently nearly eliminated serogroup A infection in vaccinated populations. However, infection by the chemically similar serogroup X remained, indicating a lack of conferred cross-protection from serogroup A and raising the question as to why this is the case.

This study finds key conformational differences between serogroup A and X, rationalizing the lack of cross-protection and adding to the call for current multivalent vaccines to be extended to include a serogroup X component. For serogroup A, the partial O-acetylation on the bacterial CPS is known to be important for antibody recognition but without a clear understanding of any potential structural changes. This

study reveals a chain-straightening effect of the O-acetyl moiety, which provides a rationale for the reported immunogenic differences between O-acetylated and de-O-acetylated chains. The finding supports the required O-acetylation of serogroup A vaccines by vaccine regulatory agencies and provides further understanding of meningococcal CPS conformations.

Conformations of Neisseria meningitidis *serogroup A and X polysaccharides: the effects of chain length and O-acetylation*

Jason Hlozek[a], Michelle M. Kuttel[b], Neil Ravenscroft[a]*

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] Department of Computer Science, University of Cape Town, Rondebosch 7701, South Africa

* Corresponding author: Tel: +27 21 650 4354; E-mail address: neil.ravenscroft@uct.ac.za (N. Ravenscroft)

Supplementary Data: Supplementary Figure S2.1

2.2 Abstract

Neisseria meningitidis is a major cause of bacterial meningitis worldwide especially in Africa. The capsular polysaccharide (CPS) is the main virulence factor and the target antigen for polysaccharide and conjugate vaccines. The high burden of serogroup A disease in the Meningitis Belt of sub-Saharan Africa led to the introduction of MenAfriVac®, which has successfully reduced the number of cases of group A disease. However, several outbreaks caused by other serogroups have been reported, including those due to serogroup X. The capsular polysaccharides of serogroups A and X are both homopolymers of amino sugars (α-D-ManNAc and α-D-GlcNAc) containing phosphodiester linkages at C-6 and C-4, respectively. The similarity of the primary structures of the two polysaccharides suggests that serogroup A vaccination may provide cross-protection against serogroup X disease. Molecular dynamics simulations of a series of serogroup A and X oligosaccharides reveal that the MenA CPS behaves as a flexible random coil which becomes less conformationally defined as the length increases, whereas serogroup X forms a more stable regular helical structure. The presence of the MenX helix is supported by NMR analysis; it has four residues per turn and becomes more stable as the chain length increases. Licensed MenA vaccines are largely O-acetylated at C-3: simulations show that these O-acetyl groups are highly solvent exposed and their presence favors more extended conformations compared to the more compact conformations of MenA without O-acetylation. These findings may have implications for the design of optimal conjugate vaccines.

Keywords: Neisseria meningitidis, Capsular polysaccharide, Molecular modeling, Antigen conformation

2.3 Introduction

The encapsulated bacterium *Neisseria meningitidis* is the causative agent of meningococcal disease, which results in meningitis, septicemia, and pneumonia. Humans are the only known host of *N. meningitidis*, which asymptomatically inhabits the nasopharynx of approximately 10% of the general population during non-epidemic periods and with varying serogroup distribution by geographic region.[37] Meningococcus is easily spread through respiratory droplets; thereafter infection occurs if the bacteria cross the mucosal membranes into the blood stream.[38] If left untreated, meningococcal disease has a high mortality rate.[39] Meningitis epidemics occur regularly on the African continent in an area extending from Senegal to Ethiopia, termed "the Meningitis Belt".[40, 41]

The bacterial capsular polysaccharide (CPS) of meningococcus is the main virulence factor and vaccination with the CPS is an effective method of limiting the spread of the disease.[42] The first successful vaccines against *N. meningitidis* infection were polysaccharide vaccines developed in the 1970s[43, 44] which are being replaced by the more immunogenic polysaccharide-protein tetravalent conjugate vaccines (Menactra, Menveo, and Nimenrix) that are more effective in children.[45]

The *N. meningitidis* CPS is used to categorize the bacteria into twelve serogroups, of which six (A, B, C, W, X and Y) cause virtually all meningococcal disease worldwide. Prior to vaccination, over 90% of meningococcal disease in the Meningitis Belt was caused by serogroup A (MenA). After introduction of a cost-effective MenA conjugate vaccine (MenAfriVac®, developed by the Meningitis Vaccine Project)[46] in 2010, MenA infections have been virtually eliminated in vaccinated populations.[47-50] However,

recent outbreaks of serogroup X (MenX) disease[42, 51] suggest the possibility of replacement of serogroup A by MenX.[48-52] This has prompted the development of a new pentavalent conjugate vaccine (NmCV-5) against serogroups A, C, Y, W, and X.[53]

The CPS of the meningococcal serogroups fall into three structurally related pairs: CPSs of serogroup A and X are phosphodiester-linked homopolymers of hexose amino sugars, serogroup B and C are homopolymers of sialic acid, and serogroup Y and W are polymers of hexose-sialic acid, as listed below.

A: →6)-α-D-Man*p*NAc(3/4OAc)-(1→OPO_3 →

X: →4)-α-D-Glc*p*NAc-(1→OPO_3 →

B: →8)-α-D-Neu*p*NAc-(2→

C: →9)-α-D-Neu*p*NAc(7/8OAc)-(2→

Y: →6)-α-D-Glc*p*-(1→4)-α-D-Neu*p*NAc(7/9OAc)-(2→

W: →6)-α-D-Gal*p*-(1→4)-α-D-Neu*p*NAc(7/9OAc)-(2→

The focus of this study is on the MenA and MenX pair, which differ in their linkage position (4-linked in MenX and 6-linked in MenA) and are epimeric at the C-2 position: the N-acetyl group is axial in MenA (*manno*) and equatorial in MenX (*gluco*). Furthermore, the MenA CPS is O-acetylated predominantly at C-3 (70 to 95%) with low levels at C-4. O-acetylation is considered to be important for immunogenicity[54] and WHO guidelines recommend that MenA vaccines have at least 61.5% O-acetylation.[55]

Oligosaccharide length is an important consideration for the development of synthetic conjugate vaccines:[56] the oligosaccharide used must be sufficiently long to be representative of the CPS conformation, but short enough for synthesis to be practical.

For the synthetic MenX vaccine in particular, it is important to determine the minimal epitope of the saccharide required for immunological activity against the MenX CPS.[57] However, there is limited data on the oligosaccharide length necessary for an immunogenic MenA and MenX glycoconjugate. A trimer of a carba-sugar analogue of MenA elicited a poor immune response in mice in comparison to oligosaccharides of 6 to 15 RU.[58] However, a MenA tetramer conjugate was shown to be antigenic in a competitive inhibition assay with rabbit antisera, although immunogenicity is still to be proven.[59] Fiebig et. al. showed that an average degree of polymerization of 10 units (avDP10) MenA conjugate had similar immunogenicity in mice as compared to the avDP15 standard.[60] For MenX, Morelli et al. synthesized monomer, dimer, and trimer glycoconjugates and found that only the trimer was able to induce IgG antibodies against MenX CPS, however, this was at low levels in comparison to the avDP15 MenX control that induced a large antibody response.[57] Although many variables affect the immunogenicity of conjugate vaccines, it is not clear what the minimum repeating unit (RU) length should be for MenA and MenX oligosaccharide-based vaccines.

Vaccine efficacy data indicates that the chemical structural similarity of bacterial CPSs alone is not able to reliably predict cross-protection between closely related strains.[26, 61, 62] Small changes in a glycan structure can lead to large changes in conformation and dynamics and thus affect antigen binding and immunogenicity. Experimental techniques such as X-ray crystallography and nuclear magnetic resonance (NMR) spectroscopy are only able to provide partial insights into glycan conformation,[63, 64] whereas computational methods can provide key theoretical insights into polysaccharide conformation. In particular, molecular dynamics (MD) simulations

provide atomistic information on dynamic polysaccharide motion in solution on the nanosecond timescale that is inaccessible to physical experiments. MD is a powerful predictor of glycan conformation which can be used to compare related antigenic glycans: our recent work on *N. meningitidis* serogroups Y and W demonstrated key conformational differences that provide a rationale for the unusual cross-reactivity observed.[27]

Here we employ MD simulations to explore the conformational preferences of MenA and MenX oligosaccharides. Our established methodology[25, 65] employs an incremental approach whereby we first simulate the respective disaccharide repeating units in order to determine the preferred individual linkage conformations. As saccharide rings are relatively rigid structures, the conformations of oligosaccharide chains are primarily determined by the torsion angles in these glycosidic linkages. The preferred torsion values are then used to build longer oligomers for MD simulation in order to generate predictions of the conformation of the CPS. We simulate 2, 6, and 10 repeating units (RU) to determine the effect of chain length, considered an important factor for the immunogenicity of glycoconjugate vaccines.[66, 67] Evaluation of predicted structures is primarily performed by cluster analysis which provides insight into the relative preference of the polymer for particular conformations. The conformational clusters can be compared between structurally similar repeating units of different saccharides to determine the extent of the role that CPS conformation plays in explaining observed immunological cross-reactivity. We use NMR NOESY correlations as corroboration for the predicted conformations. Lastly, we model MenA with 3-O-acetylation to observe the effect of this substituent on CPS conformation.

2.4 Results and discussion

In the discussion below, we first present the conformations of de-O-acetylated MenA, then MenX, and finally the effect of 3-O-acetylation on MenA. For each serogroup, we first discuss the linkage conformations [->6)-α-D-Man*p*NAc-(1->OPO_3-> for MenA and ->4)-α-D-Glc*p*NAc-(1->OPO_3-> for MenX] and then move on to conformational analysis of the 2 RU, 6 RU, and 10 RU oligomers. In the case of MenX, NMR evidence is used to corroborate our findings.

2.4.1 MenA CPS backbone conformation

The flexible α(1→6) linkage in MenA is described by five torsion angles: ϕ, ψ, ω, ε and χ (shown on a representative disaccharide linkage in Fig. 2.1a and 2.1b). The torsion angle time series from the 400 ns simulation of 6 RU of MenA (Fig. 2.1c) shows that these torsion angles vary in flexibility. The ϕ and ε torsion angles are the least flexible, each having a single conformation with average values of ϕ = 341° and ε = 179°, respectively. The χ primary alcohol torsion angle is constrained, with only brief transitions away from the primary *gg* conformation (χ = 174°). In contrast, the two C-O-P-O torsion angles (ψ and ω) are much more flexible and have strongly correlated motion in the 6 RU MenA simulation with transitions between two main conformations: a primary conformation with ψ, ω ≈ 77°,77° and a secondary conformation with ψ, ω ≈ 273°, 282° (Fig. 2.1c). The primary linkage torsion angle orients the adjacent sugar ring in an extended conformation (Fig. 2.1a), whereas the secondary conformation results in a hairpin bend conformation around the phosphodiester linkage (Fig. 2.1b).

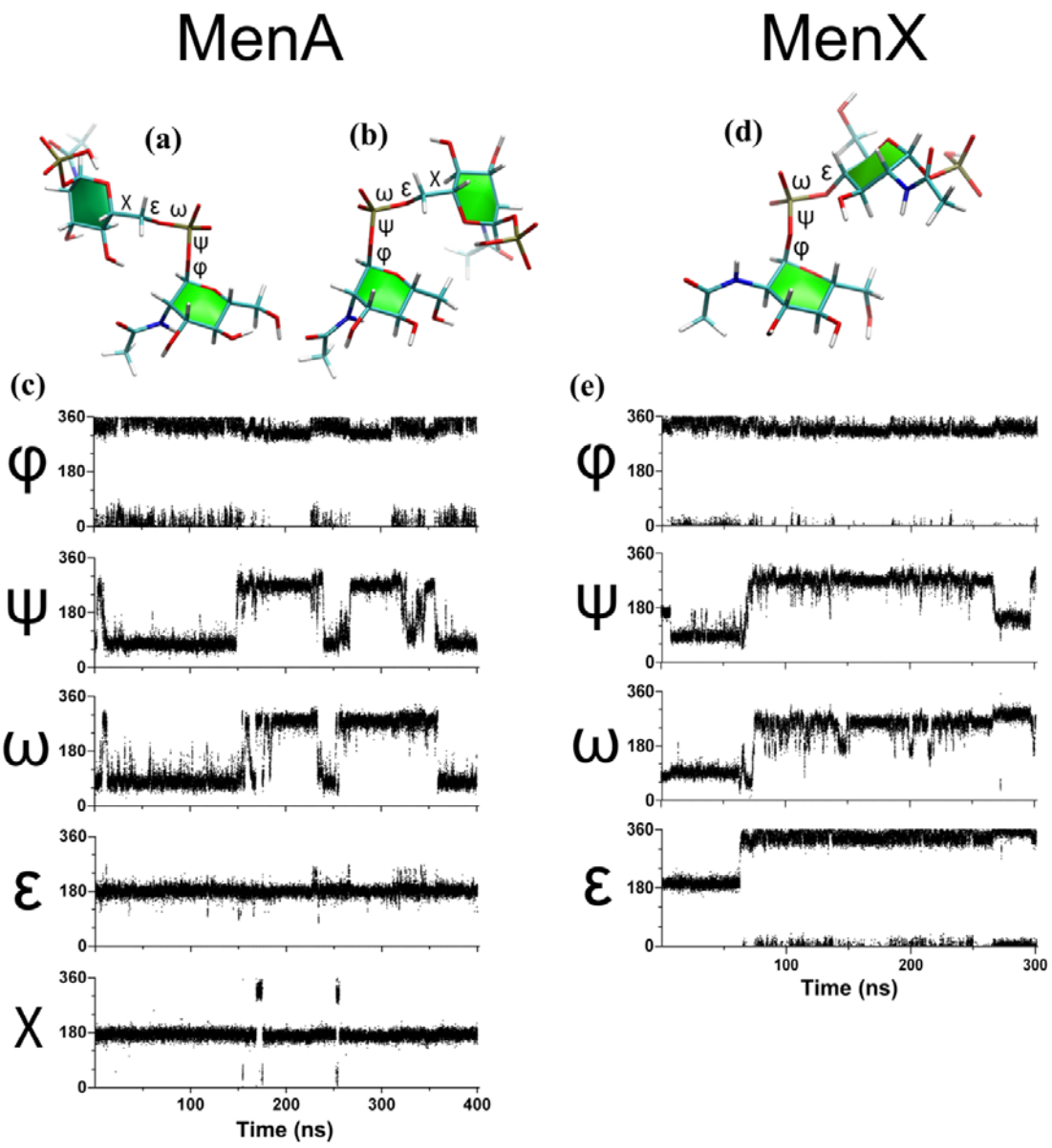


Figure 2.1. Torsion angles for the α(1→6) linkage in MenA (left column) and the α(1→4) linkage in MenX (right column). For MenA, there are two primary conformations: (a) an extended linkage conformation with φ, ψ, ω, ε, χ = 334°, 77°, 77°, 181°, 152° and (b) a hairpin bend with φ, ψ, ω, ε, χ = 325°, 271°, 282°, 178°, 174°. The corresponding torsion angle time series plots for the central linkage in a 6 RU MenA strand over the 400 ns simulation trajectory are shown in (c). The four bond linkage in MenX has a single preferred conformation with torsion angles of φ, ψ, ω, ε = 333°, 271°, 262°, 181° (d). The corresponding torsion angle time series plots for the central linkage in a 6 RU MenX strand over the 300 ns simulation trajectory are shown in (e).

As conformational changes of the α(1→6) linkage occur predominantly due to transitions in ψ and ω, we simplify the conformational preferences of the α(1→6) linkage in our simulations of 2 RU, 6 RU and 10 RU oligosaccharide strands with plots of the ψ vs ω torsion angles only. The glycosidic linkage in the 2 RU strand is predominantly in the hairpin bend conformation (ψ, ω ≈ 274°, 283°, Fig. 2.2a), with a second major population of the extended conformation (ψ, ω ≈ 77°,78°). Upon increasing the length of the oligosaccharide strand, the conformational preferences shift so that 6 RU (Fig. 2.2b) has an approximately equal weighting of hairpin bends and extended conformations, as well as a range of intermediate conformations. For the 10 RU simulation (Fig. 2.2c), the extended conformation population increases, as does the population of intermediate conformations. As repeating unit length changes, the conformational preferences of the MenA α(1→6) linkages change, therefore, the conformations of very short chains are not likely to be representative of the behavior of the native MenA CPS.

Our simulations reveal that the MenA CPS is a highly dynamic random coil with no regular conformational epitope. Clustering of the last 250 ns of the 400 ns MenA 6 RU simulation reveals a wide range of conformational families, as shown in Fig. 2.3. The 6 RU chain switches dynamically between compact (Fig. 2.3, A1, 33%), extended (Fig. 2.3, A2, 16%) and a variety of intermediate conformations (Fig. 2.3, A3, A4 and A5). This behavior of the MenA oligosaccharide correlates with reported ^{3}J ^{31}P-^{13}C coupling values that suggest a preference for extended anti conformations for the ϕ and ε torsion values.[68]

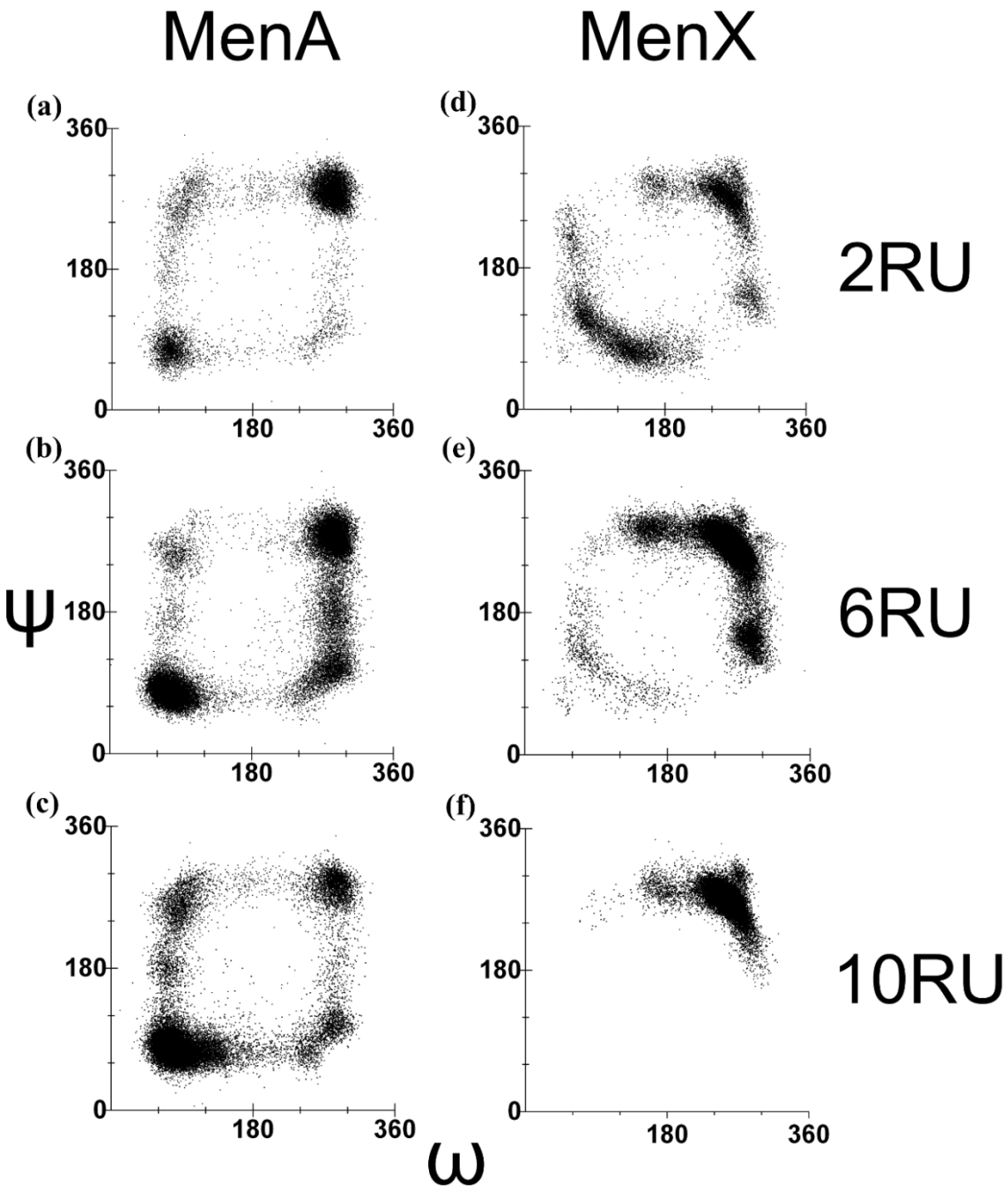


Figure 2.2. Scatter plots of the glycosidic linkage ψ, ω time series in the last 250 ns of the simulation trajectories for the α(1→6) linkages in MenA (left column) and the last 150 ns for the α(1→4) linkages in MenX (right column). For MenA, the range of motion of the glycosidic linkages increases as the chains grow in length from (a) 2 RU, to (b) 6 RU, and (c) 10 RU. In contrast, for MenX the glycosidic linkages decrease in mobility from (d) 2 RU, to (e) 6 RU, and (f) 10 RU. (Note that, for both 6 RU and 10 RU, the plots for the three central linkages in the strands are superimposed.)

2.4.2 MenX conformation

The α(1→4) linkage in MenX is described by four torsion angles: ϕ, ψ, ω and ε (shown on a representative disaccharide linkage in Fig. 2.1d). The torsion angle time series from the 250 ns simulation of 6 RU of MenX (Fig. 2.1e) shows that, as for MenA, these torsion angles vary in flexibility. This simulation was initialized with the dominant conformation from the 2 RU simulation: ϕ, ψ, ω, ε = 355°, 70°, 70°, 185°, but after equilibration for 80 ns, the ψ, ω, ε torsion angles rotated to a stable conformation of this linkage with average values of ϕ, ψ, ω, ε ≈ 320°, 270°, 265°, 340°. The ϕ torsion angle alone remained near the initial conformation. The fact that a significant conformational transition for the 6 RU strand took place only after 80 ns of equilibration illustrates the need to ensure that polysaccharide simulations are long enough to ensure thorough conformational sampling of the torsion angles; we have noted similar long equilibration times were required for simulations of 6 RU of the MenW saccharide.[27]

In contrast to MenA, the conformation of the glycosidic linkage in MenX becomes more stable as the chain length increases from 2 RU to 6 RU and 10 RU, as is clear from the scatterplots of the ψ vs ω torsion angles for the last 150 ns of MenX simulations shown in Fig. 2.2 (right column). For the 2 RU strand (Fig. 2.2d), the glycosidic linkage rotates regularly between two dominant conformations, with ψ, ω ≈ 95°, 100° and ψ, ω ≈ 270°, 265°, respectively. Upon increasing the chain length to 6 RU, the range of motion of the middle linkage (Fig. 2.2e) is clearly more constrained, showing a clear preference for the ψ, ω ≈ 270°, 265° conformation. The 10 RU simulation was begun from this dominant conformation and is even more constrained, with no rotations to

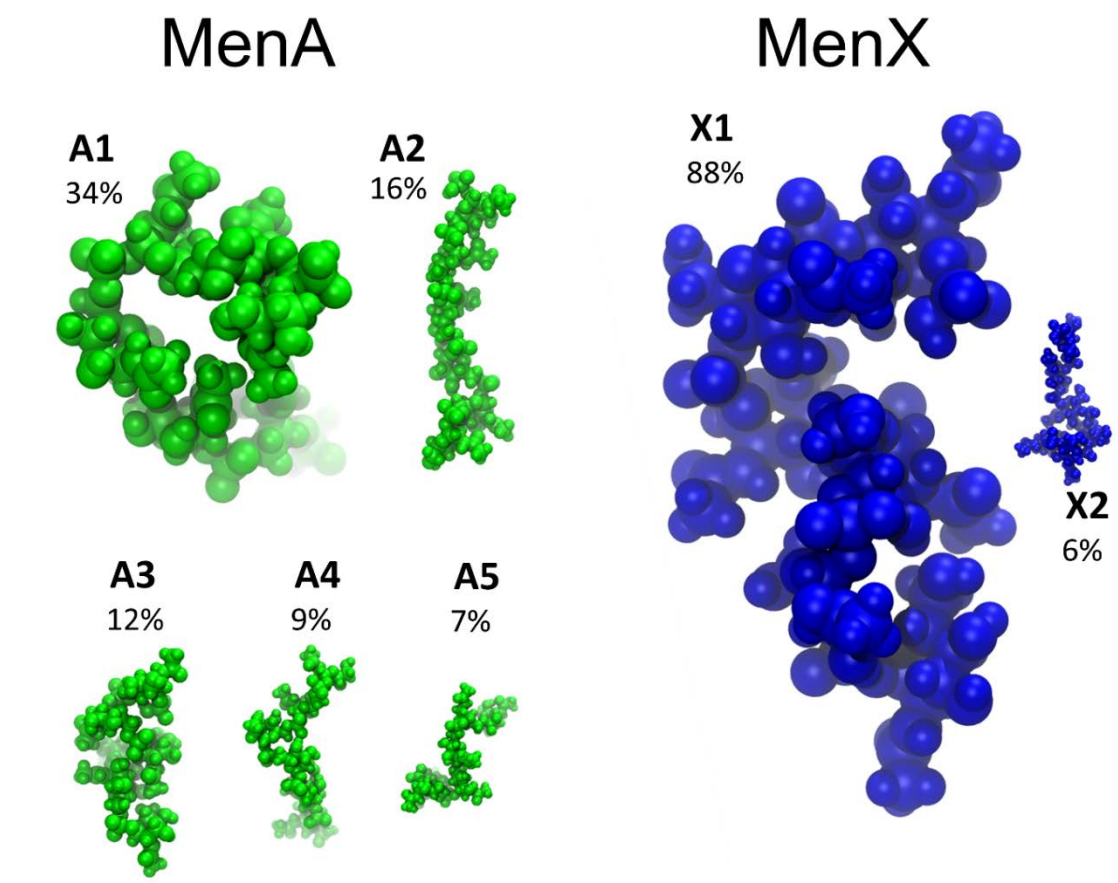


Figure 2.3. Conformational families and relative populations for 6 RU of MenA (green, left) and MenX (blue, right).

ψ, ω ≈ 95°, 100° throughout the course of the 250 ns run (Fig. 2.2f). This conformational dependence of chain length provides a rationalization for the finding that oligomers longer than 3 RU are required for immunological activity against the MenX CPS.[57]

Clustering of the MenX 6 RU trajectory conformations show that the dominant stable glycosidic linkage conformation (ϕ, ψ, ω, ε ≈ 320°, 270°, 265°, 340°) corresponds to a helix with four residues per turn (Fig. 2.3, X1). Comparison of the 6 RU and 10 RU simulations suggests that this helical conformation becomes more stable as the chain length increases: the 6 RU strand was in a helical conformation for 88% of the final

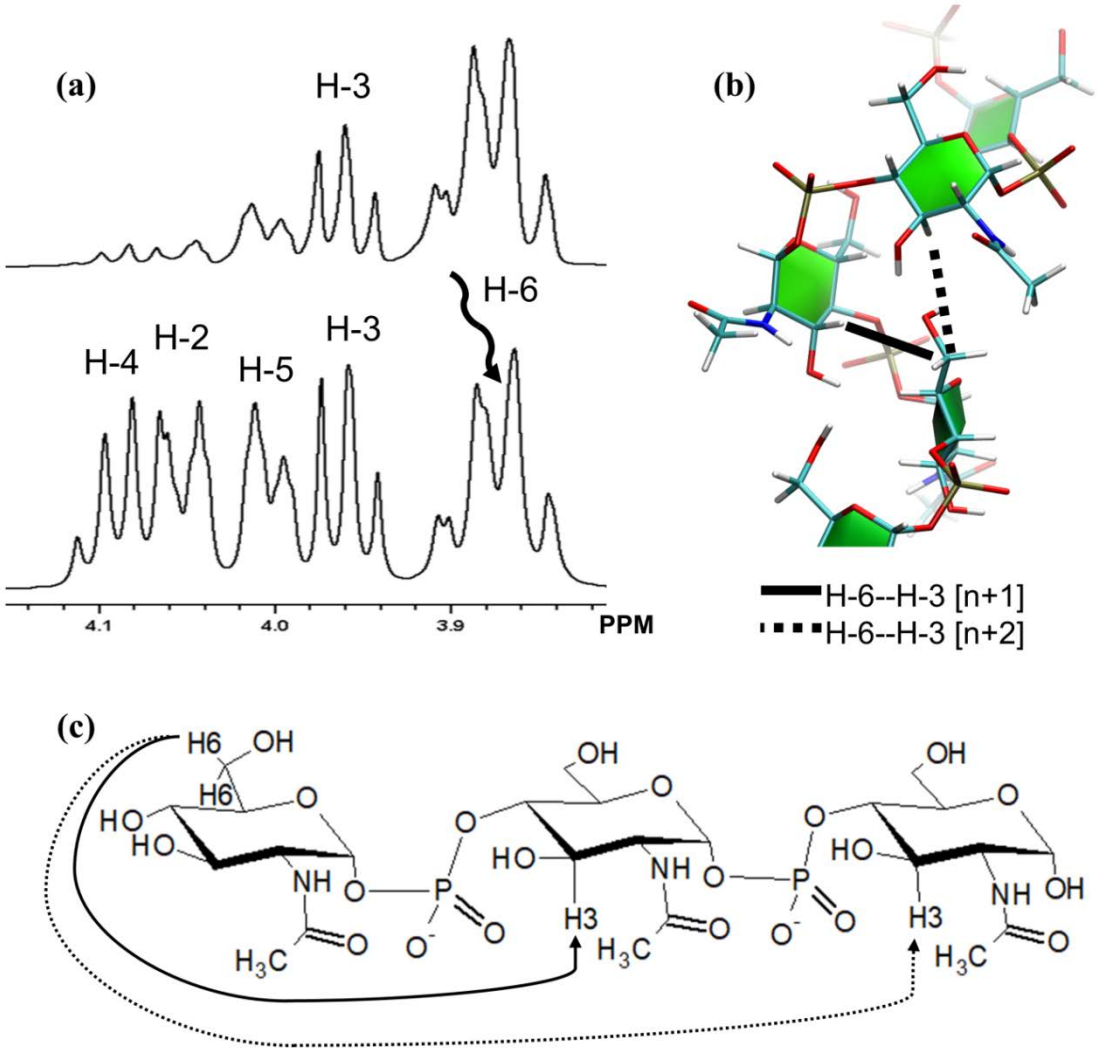


Figure 2.4: (a) Labelled expansion overlay of the 1D NOESY and 1H spectra for MenX obtained from selective irradiation of H-6, showing a strong correlation to H-3. The double correlation from H-6 to H-3 of one- and two RU along the polysaccharide is shown on the 3D model (b) as well as diagrammatically (c).

250 ns of the 6 RU simulation, whereas the 10 RU strand showed a helical conformation throughout the 250 ns run.

This hypothesis of a helical conformation of the MenX CPS is supported by key NMR NOEs obtained from 1D NOESY experiments. Differentiating between NOE correlations arising from intra-residue and inter-residue interactions for a homopolymer such as MenX is challenging because the atoms H-1 to H-6 are in the same chemical

environment for every residue. A NOESY correlation between a proton pair will have an identical chemical shift whether the correlation occurred due to intra-residue or inter-residue magnetization transfer and these are indistinguishable by NMR. Therefore, an inter-residue NOE can only be assumed where an intra-residue NOE is not possible.

A selective 1D NOESY experiment with irradiation of H-6 showed a strong correlation to H-3, indicating that H-6 and H-3 are in close proximity (Fig. 2.4a). As H-3 is on the opposite side of the GlcNAc ring to H-6 and NOE correlations to H-6 are weak for the remaining ring protons, this correlation must arise from close inter-residue contact. Indeed, the helical conformation predicted by our simulations brings H-6 into close proximity with H-3 from both the n+1 and n+2 RU (Fig. 2.4b and 2.4c). The NOESY correlation detected between H-6 and H-3 is thus strong evidence for a helical conformation of the MenX CPS.

Another inter-residue NOE correlation observed is also compatible with a helical conformation. A low intensity NOE correlation between H-6 and the NAc methyl detected again cannot be explained by an intra-residue NOE correlation because of the large distance between these groups (~9 Å). However, in the simulations, each NAc group was regularly in close proximity to the C-6 methyl on the sugar ring two RU along the polysaccharide chain (Suppl. Fig. S2.1).

2.4.3 Conformation of O-acetylated MenA

O-acetylation, 70 to 95% at C-3, is known to be important for the immunogenicity of MenA CPS.[54] Our 250 ns simulation of 6 RU of fully O-acetylated at C-3 MenA reveals

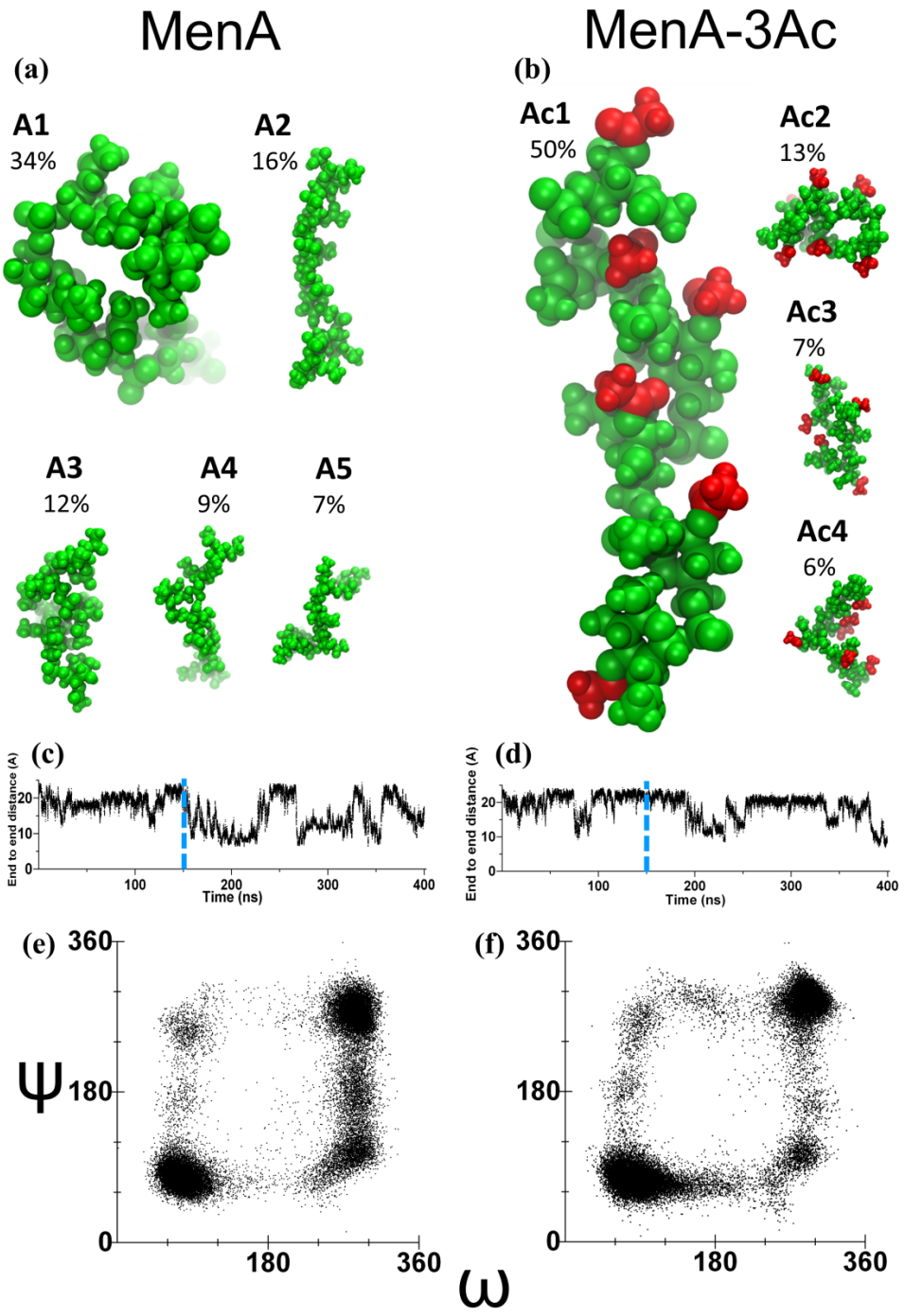


Figure 2.5: Conformational families with relative percentages for the 6 RU simulations of (a) MenA and (b) O-acetylated MenA (with acetyl groups colored in red). Time series of the end-to-end distances over the 400 ns simulations are shown for (c) MenA and (d) O-acetylated MenA; the dashed blue line indicates the point at which clustering analysis starts. Scatterplots show the ψ vs ω torsion angles for the superimposed central three linkages over the final 150 ns for (e) MenA and (f) O-acetylated MenA.

that this has an effect on the conformation of the MenA backbone. While the O-acetylated MenA has a similar family of conformations as compared to the de-O-Ac MenA, the distribution of these conformations is distinctly different. O-acetylation shifts the conformational preferences of MenA from a compact conformation (Fig. 2.5a, A1) to an extended conformation (Fig. 2.5b, Ac1): when O-acetylated, a 6 RU strand is in an extended conformation for 50% the simulation time, as opposed to only 16% for the de-OAc MenA 6 RU (Fig. 2.5a, A2). Therefore, 6 RU of O-acetylated MenA is more conformationally defined than its de-O-acetylated counterpart. This provides a rationale for the low serum bactericidal titers observed after immunization with a de-O-acetylated vaccine candidate[54] as the de-O-acetylated MenA chain does not adequately mimic the native MenA chain conformation. In contrast to MenX, a series of NOESY experiments performed on O-acetylated MenA only gave intra-residue correlations.

The increase in the extension of the 6 RU chain upon O-acetylation is apparent in a comparison of the time series graphs of the molecular end-to-end distance over time of the O-acetylated chain (compare Fig. 2.5c and Fig. 2.5d). The increased extension arises from a shift in the orientation of the α(1→6) linkage away from hairpin bends (ψ, ω ≈ 280°, 290°) towards a more extended conformation (ψ, ω ≈ 75°, 80°), as is clear in the scatterplot for MenA (Fig. 2.5e) and O-acetylated MenA (Fig. 2.5f). In the O-acetylated 6 RU oligomer, hairpin bend conformations persist for shorter periods of time. The O-acetylated MenA also has somewhat more frequent rotations around the χ torsion angle as compared to the de-O-acetylated chain.

A final observation is that the O-acetyl groups (colored red in Fig. 2.5b) in all conformational families are highly solvent exposed and therefore expected to make significant alterations to the hydrophilicity of the O-acetylated CPS surface as compared to the de-O-acetylated MenA. This altered surface for antibody binding together with the shift in conformational preferences on O-acetylation suggests that the de-O-acetylated MenA is likely to offer at most limited cross-protection against the O-acetylated native MenA CPS, thereby explaining why O-acetylation is important for MenA vaccines.

2.5 Conclusions

Comparison of the MenA and MenX simulations shows that the flexible five-bond α(1→6) glycosidic linkage in MenA produces a markedly more disorganized saccharide chain than the α(1→4) linkage in MenX: the 6 RU MenA oligosaccharide behaves as a random coil in aqueous solution, whereas 6 RU of the MenX CPS forms a regular helix. The helical model for MenX is corroborated by evidence from selected NMR NOESY NOEs. The conformation of the MenX oligosaccharide is length dependent, with the helical conformation becoming increasingly stable as the chain length increases from 2 RU to 10 RU. As the stable helix conformation has four residues per turn, our simulations suggest that the minimal epitope required for the immunological activity against the MenX CPS will be at least 4 RUs. The large difference in structure and conformation between MenA and MenX provides a rationale as to why no cross-protection against MenX infection from MenA vaccination is observed, thereby validating the need for a MenX vaccine.

In addition, our simulations reveal that O-acetylation has an effect on the conformation of the MenA saccharide chain: O-acetylation favors extended conformations over more compact conformations with hairpin bends of the α(1→6) glycosidic linkage. This conformational difference, together with the fact that the O-acetyls are highly solvent exposed, may explain why O-acetylation is important for MenA vaccines.

2.6 Experimental

We analyzed the MenA and MenX saccharides according to the glycosidic linkage torsion angles. For the five-bond α(1→6) phosphodiester linkage of MenA the torsion angles are here defined as: ϕ = H1-C1-O1-P, ψ = C1-O1-P-O6', ω = O1-P-O6'-C6', ε = P-O6'-C6'-C5' and χ = O6'-C6'-C5'-H5'. The four-bond α(1→4) phosphodiester linkage in MenX has torsion angles defined similarly as ϕ = H1-C1-O1-P, ψ = C1-O1-P-O4', ω = O1-P-O4'-C4', ε = P-O4'-C4'-H4'.

2.6.1 Molecular dynamics simulations

All simulations were performed with the NAMD molecular dynamics program[33] with CUDA extensions to enable calculation of long-range electrostatic forces and non-bonded forces on graphics processing units.[69] Glycans were modeled with the CHARMM36 additive force field for carbohydrates,[31, 32] incorporating ad hoc extensions to represent the phosphodiester linkages and the acetyl substituents. Water was simulated with the TIP3P water model,[70] which is compatible with the CHARMM carbohydrate force field.

The initial structures for simulations were in all cases built using our in-house software CarbBuilder,[71, 72] which employs the psfgen tool to create Protein Structure Files (PSF) for modeling with the CHARMM force field in the NAMD molecular dynamics program. Structures for the 6 RU oligosaccharides were built using the most common torsion angles from the respective 2 RU simulations, and the 10 RU simulations similarly used the most common values from the 6 RU simulations. All simulations began with 10,000 steps of standard NAMD minimization in vacuum followed by solvation in a water box using the *solvate* routine in the Visual Molecular Dynamics (VMD) software.[73] To reduce computational cost, a minimal water box size was chosen for each simulation whilst still sufficiently enveloping the respective chains: the 2 RU and 6 RU simulations used a 60 $Å^3$ periodic cubic cell (approximately 7,000 water molecules) and the 10 RU structures had 80 $Å^3$ for MenA 10 RU and 70 $Å^3$ for MenX 10 RU (16,000 and 11,000 water molecules respectively). This variation in water box size implies a difference in effective solute concentration that may be relevant in expanded study into interactions between multiple chains or of the chain with salt ions. Randomly-distributed sodium ions were added to each system using the VMD *autoionize* feature in order to electrostatically neutralize the negative charge from the phosphodiester linkages.

All MD simulations were preceded by a 122,000 step minimization and heating phase, consisting of 5 K temperature reassignments from 10 K to a maximum temperature of 310 K with 1,000 steps of minimization and 1,000 steps of MD at each temperature. The 2 RU MenA and MenX simulations and the 6 RU MenX simulation ran for 300 ns whereas, in order to allow for equilibration, the highly flexible MenA and MenA-3Ac 6

RU simulations were run for 400 ns. The 10 RU simulations are computationally expensive and so were only run for 250 ns. While this was not sufficient for full conformational equilibration, this allowed us to probe the conformational preferences of these longer strands.

For the MD, equations of motion were integrated using a Leap-Frog Verlet integrator with a step size of 1 fs and periodic boundary conditions. Simulations were performed under isothermal-isobaric (nPT) conditions at 310 K maintained using a Langevin piston barostat[33] and the Nose-Hoover thermostat implemented in NAMD, which is a combination of a Nose-Hoover constant pressure method[74] with piston fluctuation control implemented using Langevin dynamics.[75] Long-range electrostatic interactions were treated with particle mesh Ewald (PME)[76] summation using k = 0.20 A^{-1} and PME grid dimensions equal to the system periodic cell dimensions. Non-bonded interactions were truncated with a switching function applied between 12.0 and 15.0 Å. The 1-4 interactions were not scaled, in accordance with CHARMM force field recommendations.

2.6.2 Post-simulation analysis

Molecular structures were visualized with VMD,[73] where necessary using the PaperChain visualization algorithms[77] to highlight the glycan rings.

For each simulation trajectory, a time series of molecular conformations spaced 12.5 ps apart was extracted for analysis. Analysis of the ψ and ω torsion angles used only the final 150 ns of each simulation. For the 6 RU and 10 RU saccharides, the ψ and ω scatter plots combined the time series from the three middle glycosidic linkages in the

chains. End-to-end distances for the 6 RU saccharide chains were defined to exclude the terminal residues, from C-1 of residue 2 to C-4 of residue 5.

For clustering, the 6 RU simulation conformations were aligned on the middle 2 RU and then clustered into conformational families using VMD's internal *cluster* command, which applies the quality threshold algorithm.[78] The clustering metric comprised an RMSD fit of the saccharide backbone for the middle 4 RU in the 6 RU strand (where the backbone is defined as the glycan ring carbon and oxygen atoms, as well as the oxygen and phosphorous atoms in the glycosidic linkages) with a cut-off value of 3.5 Å. The first 150 ns of each simulation were treated as equilibration and not clustered.

Theoretical ^{3}J ^{31}P-^{13}C coupling values for comparison with NMR literature were calculated using a Karplus equation parameterized for ^{31}P-O-C-^{13}C torsion angles.[79]

2.6.3 NMR analysis

MenA and MenX polysaccharide samples (~10 mg) were lyophilized and exchanged twice with 99.9% deuterium oxide, then dissolved in 600 µL of D2O and introduced into a 5 mm NMR tube for data acquisition. MenA and MenX polysaccharide samples (~10 mg) were lyophilized and exchanged twice with 99.9% deuterium oxide (D_2O), then dissolved in 600 µL of D_2O and introduced into a 5 mm NMR tube for data acquisition. ^{1}H, 1D selective NOESY (selnogp) and 2D NOESY (noesyphpr) spectra were recorded using a range of mixing times (100 to 300 ms for MenA and 100 to 400 ms for MenX, D1 of 2s) on a Bruker Advance III 600 MHz NMR spectrometer. For the 1D NOESY, the isolated H6 peak was selected by integration and subsequently irradiated with a band width of 100 Hz. The spectrometer was equipped with a BBO Prodigy

cryoprobe set at a temperature of 303 K and NMR spectra were processed using standard Bruker software (Topspin 3.2).

Declarations of interest

N. R. provides NMR analysis to PATH's projects including the polyvalent meningococcal conjugate vaccine in development.

Acknowledgments

Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za. The authors thank the Serum Institute of Indian Pvt Ltd (SIIPL) for providing the MenA and MenX polysaccharides that were prepared as part of a collaboration between SIIPL and PATH to develop a polyvalent meningococcal conjugate vaccine. The authors thank the South African National Research Foundation for NMR equipment funding (NRF Grant 86038).

2.7 Supplementary information

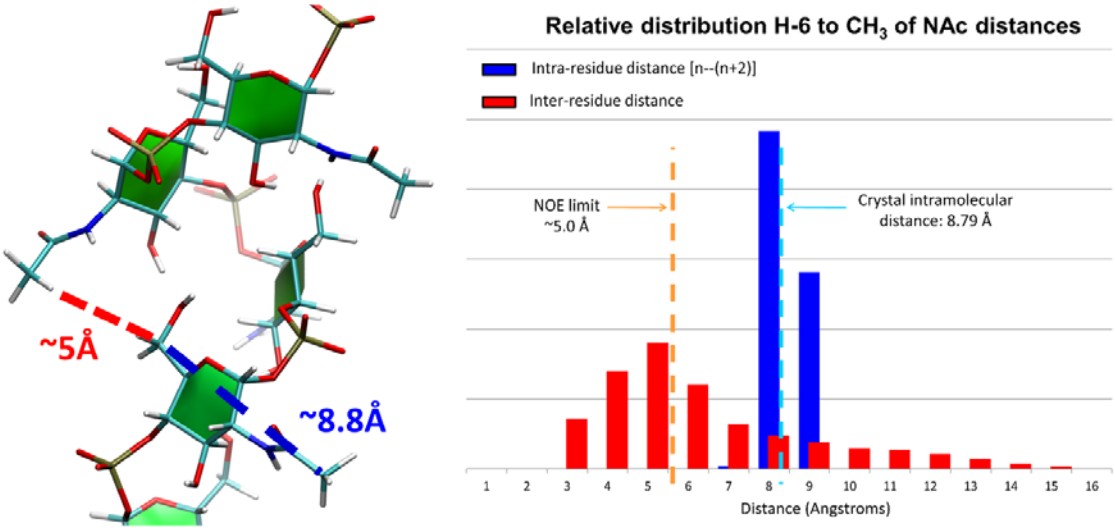


Figure S2.1: Left panel: 3D helical model of MenX showing H-6 to NAc methyl group distances that illustrates the inter-residue NOE observed (left panel). Right panel: The relatively rigid intra-residue H-6 and NAc methyl group distance is consistently longer than the ~5 Å limit for NOE correlations (blue). However, the inter-residue distance of H-6 and the NAc methyl group of the residue two RU away (red) is regularly within this NOE distance limit.

Chapter 3

Analysis of a chemically-stable mimic of *N. meningitidis* serogroup A

3.1 Foreword

The following article "*Comparison of the conformation of* Neisseria meningitidis *serogroup A CPS with a carba-analogue: implications for vaccine development*" extends the study of meningococcal serogroup A. The CPS of serogroup A is chemically-labile, which limits shelf-life and requires a two-vial presentation that increases storage and transportation costs. A proposed solution is to stabilize the glycosidic linkages by synthetically replacing the electron-withdrawing ring-oxygen with a methylene group, however, it is unclear how this might impact overall polysaccharide conformation.

The study finds key conformational differences between the native CPS and the carba-analogue chain, suggesting limited cross-protection from the proposed replacement stable analogue. This publication also notes significant differences between the conformations of short glycan chains and the corresponding longer chains. Indeed, this variance in immune response has been observed in clinical study of MenA antigens where no bactericidal activity was detected in mice immunized with a 1 RU or 2 RU Carba-MenA glycoconjugate, while a moderate response was observed in mice

immunized with a 3 RU Carba-MenA glycoconjugate.[58] However, the same study found significant antibody responses for longer 6 RU and 15 RU native MenA glycoconjugates. These findings highlight a need for caution when selecting antigen chain lengths to represent the native glycan polymer.

Modeling the conformations of Neisseria meningitidis *serogroup A CPS and a carba-analogue: implications for vaccine development.*

Jason Hlozek[a], Neil Ravenscroft[a] and Michelle M. Kuttel[b]*

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] Department of Computer Science, University of Cape Town, Rondebosch 7701, South Africa

* Corresponding author: Tel: +27 21 650 5107; E-mail address: mkuttel@cs.uct.ac.za (M. M. Kuttel)

3.2 Abstract

Neisseria meningitidis is a major cause of bacterial meningitis worldwide, especially in Africa. The capsular polysaccharide is the main virulence factor and the target antigen for polysaccharide- and conjugate vaccines. Three tetravalent conjugate vaccines against serogroups A, C, Y and W have been licensed and the monovalent MenAfriVac® was introduced to address the high burden of serogroup A disease in the Meningitis Belt of sub-Saharan Africa. Three of these four vaccines are lyophilized due to the instability of the serogroup A antigen (MenA) in aqueous solution, resulting in a two-vial presentation with concomitant additional costs for storage and distribution. Replacement of the saccharide ring oxygen with a methylene group is a promising approach to preparing a stable oligosaccharide MenA analogue (Carba-MenA) vaccine suitable for a liquid formulation. However, to be effective, Carba-MenA must elicit an immune response that is cross-reactive to the native MenA. Here we employ microsecond molecular dynamics simulations of ten repeats of MenA and Carba-MenA to establish that there are significant differences in the conformation and dynamics of these antigens in solution. Carba-MenA has a more random extended, conformation than MenA; MenA has a significant population of compact S-bend conformations that are absent in the analogue. We also find that the disaccharides are poor models of the conformational behavior of longer chains. This information is relevant for the rational design of optimal analogues for conjugate vaccines.

Keywords: Neisseria meningitidis, Serogroup A, capsular polysaccharide, carba-analogue, molecular modeling, antigen conformation, vaccine

3.3 Introduction

Meningitis and septicemia caused by the encapsulated bacterium *Neisseria meningitidis* has significant morbidity and mortality in children and young adults worldwide. There are twelve known serogroups of *N. meningitidis*, six of which - A, B, C, W, X, and Y - are associated with epidemic infection. Meningococcal epidemics have occurred regularly in the "Meningitis Belt" region of sub-Saharan Africa, reaching from Senegal to Ethiopia.[40] The meningococcal capsular polysaccharide (CPS) is the primary virulence factor and vaccine target: vaccination with saccharide-protein conjugates has reduced the burden of meningococcal disease worldwide.[42]

Prior to vaccination campaigns in 2010, meningococcal epidemics reached incidence rates of 1,000 cases per 100,000 population, of which over 90% were due to serogroup A (MenA) infection.[80] MenA produces a homopolymeric CPS comprising an α(1→6)-linked phosphodiester N-acetyl-D-mannosamine (Fig. 3.1a) with O-acetylation: →6)-α-D-Man*p*NAc(3/4OAc)-(1→OPO_3 →. The introduction of an affordable serogroup A vaccine (MenAfriVac®) in 2010[46] has virtually eliminated MenA infection in vaccinated populations,[47-50] demonstrating the efficacy of vaccination against meningococcal infection. The licensed tetravalent meningococcal conjugate vaccines (Menactra®, Menveo™ and Nimenrix®) contain MenA together with serogroups C, Y and W. The instability of the MenA antigen in aqueous solution, attributed to the presence of the axial N-acetyl group at C2 of ManNAc,[81] means that three of these four licensed vaccines contain MenA in the more stable lyophilized form. This requires a two-vial presentation of the vaccine, with extra costs in storage and distribution.

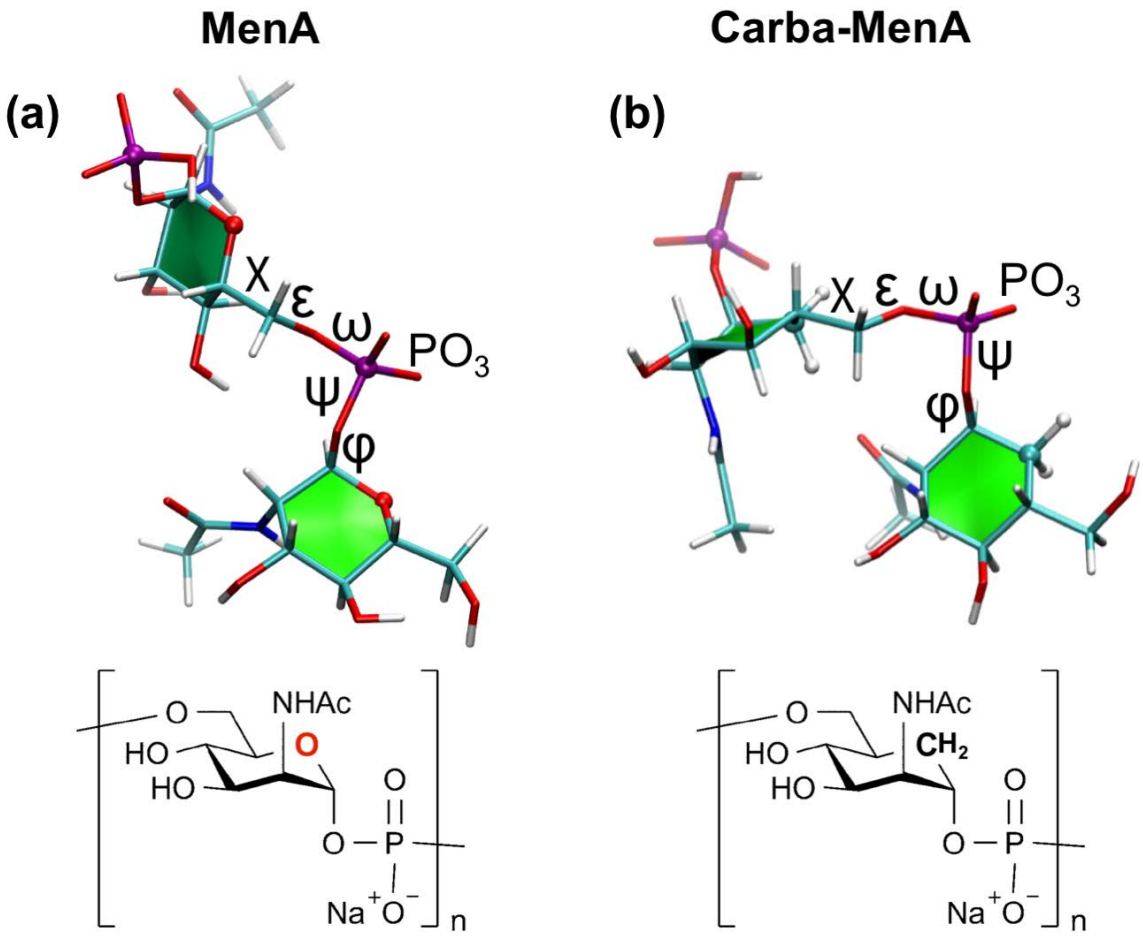


Figure 3.1: Representative disaccharide repeating units of the CPSs of (a) MenA and (b) Carba-MenA, showing the torsion angles defining the linkage orientations and the location of the phosphate. The repeating unit structures are shown below: the Carba-MenA analogue substitutes the N-acetyl-D-mannosamine ring oxygen with a methylene group.

Several approaches have been taken to the synthesis of stable MenA CPS analogues suitable for a liquid formulation: C-phosphonates,[82, 83] a carba-sugar and, more recently, a triazole analogue.[84-86] The most promising of these (here termed Carba-MenA) replaces the ring oxygen of N-acetyl-D-mannosamine with a methylene group – compare Fig. 3.1a with Fig. 3.1b.[87]

To be an effective antigen, an analogue must elicit an immune response that is cross-reactive with the native MenA CPS. There also is the question of chain length: for MenA, chains with more than four repeating units (RU) appear to provoke stronger antigenic responses, with optimal immunogenicity predicted for 5-7 RU.[56] Harale et. al. reported an antigenic MenA tetramer conjugate, though immunogenicity is still to be proven.[59]

The precise mechanisms involved in antigen cross-protection remain unclear, but it is generally understood that structural similarity is not sufficient to ensure cross-protection for CPS antigens.[26, 61, 62] Because small alterations in saccharide structure (e.g. in a residue type, linkage position, substitution, and side chains) can produce significant changes in molecular conformation and flexibility, conformational similarity of the antigens is also expected to be necessary for antigen cross-protection.[1] Molecular modeling is playing an increasingly important role in prediction of the conformation of flexible glycans in solution, both in combination with experiment[25, 88] and on its own.[27] Recent modeling studies have predicted significant conformational effects resulting from O-acetylation in MenA[1] and the typhoid CPS,[89] as well as from branching as described for the O-polysaccharides of *Klebsiella pneumoniae*[90] and *Moraxella catarrhalis.*[91] In particular, comparison of the conformation and dynamics of MenA and Carba-MenA antigens is expected to provide insight into the viability of replacing the unstable MenA with a more stable analogue. Initial modeling studies on short antigen fragments show promise: quantum mechanical calculations for 1 RU predicted that the $^4C_1$ chair configuration of the mannose ring dominates in both the MenA phosphomethyl-monosaccharide and the Carba-MenA analogue.[92] More recent

molecular dynamics simulations for these 1 RU fragments indicated similar conformations, on the basis of which a helical conformation was predicted for both the MenA and Carba-MenA polymers.[93] However, it is unclear whether a single monomer is an appropriate model for the corresponding longer oligosaccharide: these results are at variance with our simulations of a long MenA strand (10 RU), which showed the saccharide chain to be highly flexible, with no evidence of stable helical conformations.[1]

Therefore, there remain important conformational questions to be answered for the MenA CPS and its Carba-MenA analogue, as follows. Does methylene substitution of the mannosamine ring result in a significant difference in the conformation of MenA as compared to Carba-MenA? Do MenA and Carba-MenA chains form stable helices in solution? What is the optimum chain length for the MenA antigen?

To shed light on these questions, we extend our previous simulations of MenA (2 RU and 10 RU) to the microsecond time scale and compare these to new Carba-MenA simulations (2 RU and 10 RU). Comparison of the conformations and dynamic motion for the chains highlights the similarities and differences in conformation and flexibility of the MenA and Carba-MenA antigens, as discussed below.

3.4 Results and discussion

We begin with analysis and comparison of the 10 RU simulations for the MenA and Carba-MenA homopolymer chains and then compare these with the corresponding 2 RU simulations.

3.4.1 Oligosaccharide chain conformations

Our prior study indicated that the MenA CPS behaves as a flexible random coil[1] and the extension of the original 300 ns simulation to 1 μs corroborates this finding. Here we use the end-to-end distance of the 10 RU strand, *r1*, as a simple measure of conformational extension of the 10 RU saccharide (Fig. 3.2a). The time series of *r1* (specifically the distance from C1 of the reducing end to C5 of the non-reducing end of the strand) shows rapid transitions between a broad range of values from 9.1 Å to 59.1 Å (Fig. 3.2b), with a mean of 33.0 Å (standard deviation 8.5 Å). The corresponding histogram of *r1* (Fig. 3.2c) shows a normal distribution, which indicates random coil behavior. However, although flexible on a large scale, the MenA 10 RU chain does display some local order, as is revealed by an alternative distance measure, *r2*, the distance from C1 of RU4 to C5 of RU7 (Fig. 3.2a). The MenA time series for *r2* (Fig. 3.2d) also shows considerable flexibility, but alternates between two primary values*:* a population of relatively compact conformations centered on $r2 \approx 13$ Å (53%), and a slightly smaller population of extended conformations centered on $r2 \approx 22$ Å (47%). This bimodal distribution of lengths is clear in the corresponding histogram of *r1* (Fig. 3.2e) and provides evidence of local order in the 4 RU central residues in the MenA strand.

The Carba-MenA 10 RU strands also shows considerable flexibility, as evidenced by the broad range of *r1* values explored during the simulation (Fig. 3.2f). However, the range of values is wider than for MenA (Fig. 3.2g): from 5.6 Å to 59.9 Å, with a mean of 30.4 Å (standard deviation 10.9 Å). The *r2* time series for Carba-MenA (Fig. 3.2h) shows a greater difference in the central 4 RU as compared to MenA. In contrast to the bimodal

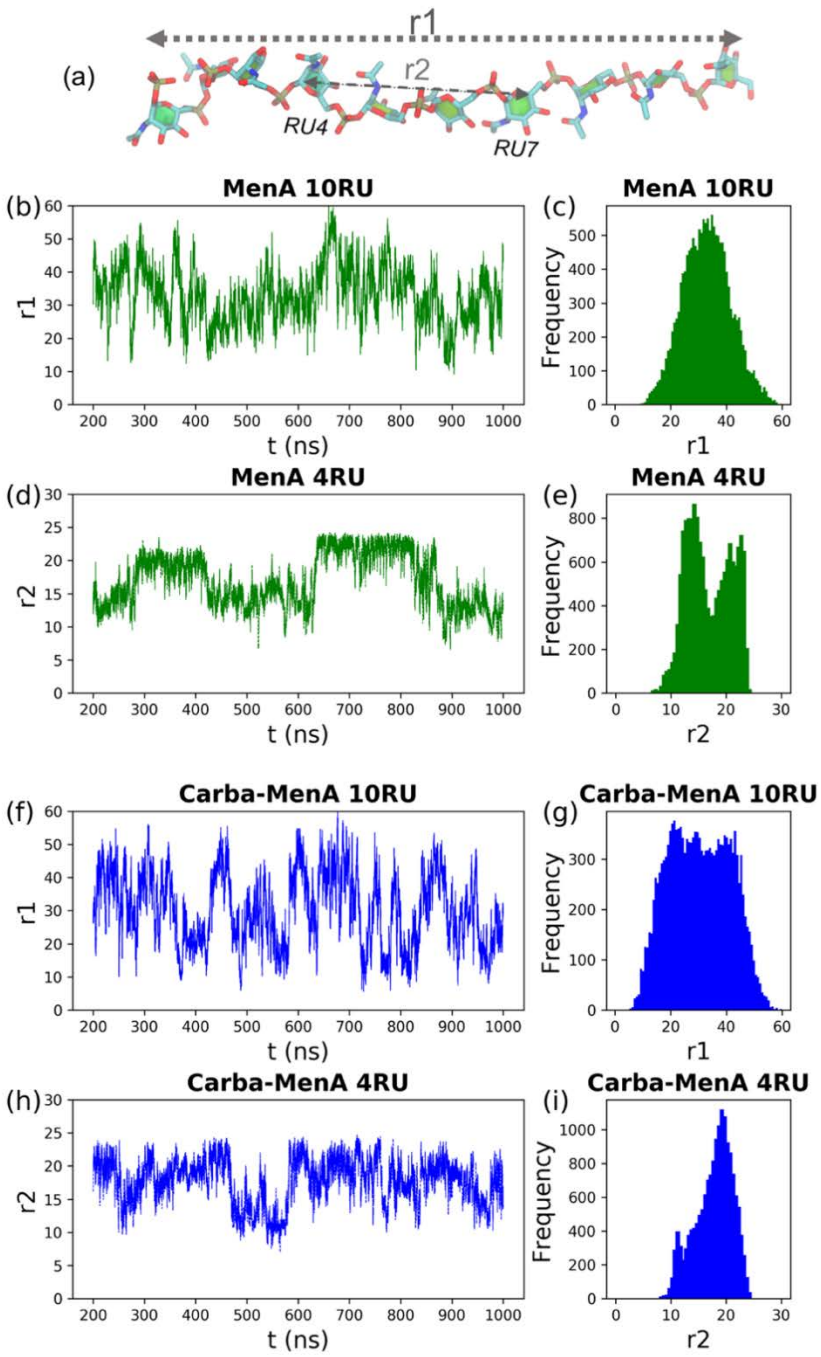


Figure 3.2: Comparison of the end-to-end distance time series for MenA and Carba-MenA, excluding the first 200 ns of simulation time. (a) We define two end-to-end distances: r1 is the distance from C1 of the reducing end to C5 of the non-reducing end of the strand; r2 is from RU4 to RU7 and thus excludes the terminal residues. For MenA: (b) r1 time series and (c) r1 histogram; (d) r2 time series and I r2 histogram. The corresponding graphs for Carba-MenA are shown in (f), (g), (h) and (i) respectively. Distances are in Å.

MenA, the distribution of *r2* for Carba-MenA is more unimodal and centered on $r \approx 19\ \text{Å}$.

The causes of this significant difference in the conformation of the central 4 RU in the MenA and Carba-MenA strands are revealed by grouping the conformations into families (Fig. 3.3). MenA has two primary conformational families of almost equivalent prevalence – Cluster A1 at 32% and Cluster A2 at 27% (Fig. 3.3a). More minor clusters (A2, A3 and others) occupy insignificant fractions of the simulation time. Cluster A1 is an extended ribbon-like conformation with $r \approx 22\ \text{Å}$ and corresponds to the second peak in the *r2* histogram. In this A1 conformation, the phosphate groups are aligned along one edge of the ribbon and the N-acetyl groups aligned along the opposite edge. The A1 conformation is stabilized by hydrogen bonds between the O4 hydroxyls and the N-acetyls on the adjacent residues. In contrast, Cluster A2 is a compact "S-bend" conformation with $r \approx 13\ \text{Å}$, corresponding to the first peak in the *r2* histogram. This conformation has a tight bend at the glycosidic linkages and is stabilized by hydrogen bonds between alternating residues, i.e. between RU4 O4 and the RU6 phosphate as well as between RU5 O4 and the RU7 phosphate.

The primary conformational cluster for the Carba-MenA analogue, cA1 (Fig. 3.3b), is an extended conformation similar to A1, with $r2 \approx 22\ \text{Å}$. This cluster dominates throughout the simulation (41%) and corresponds to the single peak in the *r2* histogram for Carba-MenA. As is the case for the A1 cluster in MenA, the cA1 cluster is stabilized by hydrogen bonds between the O4 hydroxyls and the N-acetyls on the neighboring linkages, although these hydrogen bonds occur less frequently. A much less significant secondary cluster, cA2 (17%), corresponds to a bend in the chain, with

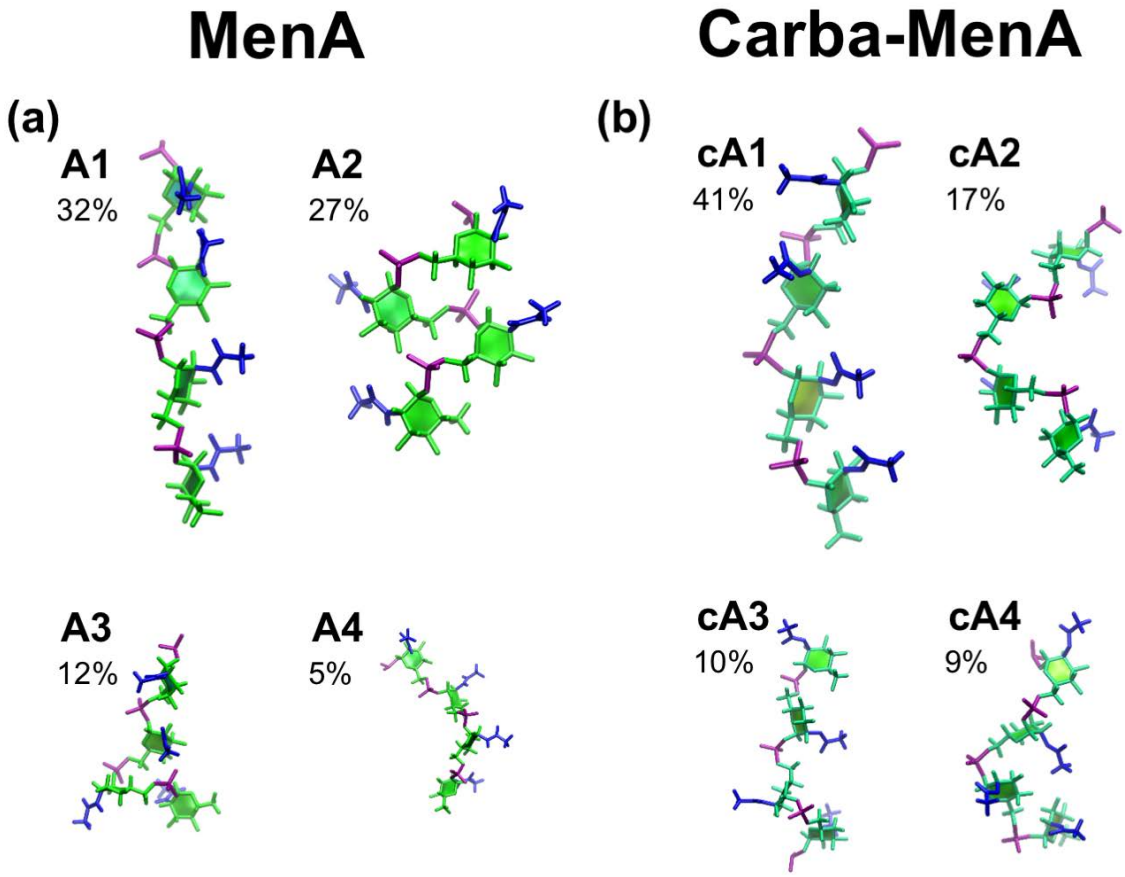


Figure 3.3: Conformational families generated by clustering analysis of the central 4 RU of the 10 RU strands: (a) MenA and (b) Carba-MenA. The relative percentage occupancy of each conformational family is indicated.

$r \approx 15$ Å. The compact "S-bend" conformation is absent for Carba-MenA, the range of smaller conformational clusters (cA3, cA4 and others) are predominantly extended conformations. The Carba-MenA strand thus exhibits a greater range of conformations and is generally more extended than MenA, especially on a local (4 RU) scale. These differences in conformation between MenA and Carba-MenA are caused by the ring methylene group substitution in Carba-MenA changing the dominant orientation of the phosphodiester glycosidic linkages, as discussed below. However, this substitution does not alter the ring conformations: both MenA and Carba-MenA maintain $^4C_1$ chair

conformations of the pyranose rings throughout the simulations, in agreement with DFT calculations for the monosaccharide repeat units.[92]

3.4.2 Glycosidic linkage conformations

The orientation of the flexible α(1->6) phosphodiester linkage in MenA is determined by the rotation about all the five bonds in the linkage, which is conveniently described by five torsion angles: ϕ, ψ, ω, ε, and χ (labelled on the representative linkages in Fig. 3.1a MenA and Fig. 3.1b Carba-MenA). These torsion angles vary in their preferred orientations and mobility. Our analysis focuses on the center linkage in the 10 RU chain (linkage RU5-RU6) as this is representative of the corresponding polymer. For this center linkage, the ϕ and ε torsion angles show little flexibility in both oligosaccharides, oscillating around a single conformation: ϕ = -15°/15° (MenA/Carba-MenA) and ε = 180° (both). The ϕ torsion angle is shifted in Carba-MenA, likely due repulsion from the hydrogens in the methylene group substituted for O5.

The remaining torsion angles are more flexible, with frequent conformational transitions over the course of the 1 μs simulations. Both the ψ and ω (C-O-P-O) torsion angles transition frequently between two main conformations in MenA and Carba-MenA: 85° and -85°. The χ (primary alcohol) torsion angle has three favored orientations: the *gg* (-60°) and the *gt* conformations (60°) dominate, with a minor *anti* (180°*)* population. Carba-MenA shows a lower *anti* population relative to MenA.

Experimental corroboration for the torsion angle conformations of the MenA CPS in solution is challenging to obtain, as the MenA CPS homopolymer produces NMR NOE signals that cannot be unambiguously assigned to either inter- or intramolecular correlations. However, the dominant extended conformation we find for the MenA

oligosaccharide correlates with reported ^{3}J ^{31}P-^{13}C coupling values that suggest a preference for extended anti-conformations of the ϕ and ε torsion angles.[68] In addition, an extended conformation is in agreement with a recently reported crystal structure of a MenA trimer in complex with a monoclonal antibody, which exhibits a linear epitope.[94]

A focus on the orientation of individual torsion angles is not appropriate for MenA/Carba-MenA, as there are a variety of rotations about the flexible ψ, ω and χ torsion angles that will produce similar overall orientations of the linkage; thus, it is difficult to draw conclusions based on the analysis of the orientation of individual bonds described by torsion angles. However, clustering the orientation of the central glycosidic linkage (RU5-RU6) in each chain into conformational families indicates the dominant orientation of this linkage throughout the simulations. Such a clustering analysis reveals significant differences between the dominant glycosidic linkage orientations for MenA and Carba-MenA, as shown in Fig. 3.4. For MenA, there is a direct correspondence between the dominant linkage orientations (L1 and L2, Fig. 3.4a) and the dominant conformations of the central 4 RU (A1 and A2, Fig. 3.3a). The central linkage of the MenA strand is split almost equally between extended conformations (L1 at 40%) and tight hairpin bend conformations (L2 at 39%). In contrast, in Carba-MenA the extended conformations dominate (Fig 3.4b, cL1 at 69%) as hairpin bends are very infrequently sampled (cL5 at 6%) and there are intermediate conformations that are not significantly populated in MenA (cL2 at 16%, cL3 at 8% and cL4 at 6%).

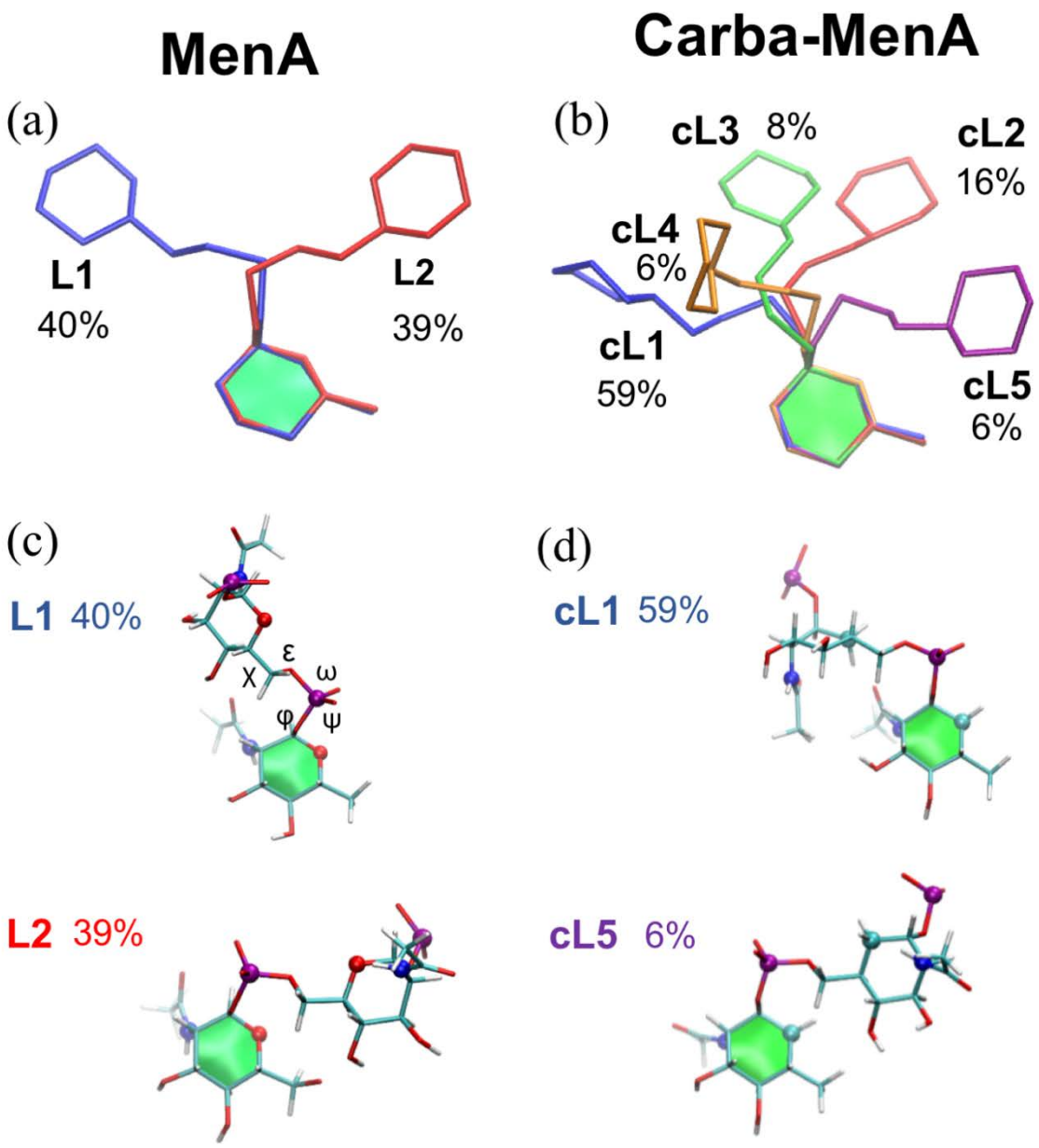


Figure 3.4: The dominant conformations of the central linkage in the 10 RU simulations of (a) MenA and (b) Carba-MenA, annotated with relative percentage occupancy over the 800 ns production run (200 ns equilibration ignored). All atom representations of conformations of the central linkage are depicted below for (c) MenA (L1 and L2) and (d) Carba-MenA (cL1 and cL5).

In the MenA L2 conformation, a tight hairpin bend conformation of the phosphodiester linkage brings the aliphatic hydrogens on C6 into close proximity with the ring oxygen on the neighboring sugar residue in MenA, as shown in Fig. 3.4c. For the Carba-MenA analogue, O5 is replaced with a methylene group, with consequent repulsive forces between the hydrophobic hydrogens in bent conformations (Fig. 3.4d, conformation cL5). Therefore, hairpin bends are less common in the Carba-MenA analogue. In addition, the relative orientation of the saccharide rings in the dominant extended conformation differs in MenA as compared to Carba-MenA: in Carba-MenA the saccharide rings are orientated perpendicular to the neighboring sugar as compared to the parallel orientation in MenA (compare conformations L1 and cL1 in Fig 3.4.).

Therefore, this significant difference in glycosidic linkage conformations between MenA and Carba-MenA can be directly linked to the altered conformational preferences of the central 4 RU, as for the 10 RU oligosaccharide chains, as follows. Although these are both flexible chains, overall MenA is more conformationally defined than Carba-MenA, with oscillations between S-bend and extended orientations of the strand. In contrast, Carba-MenA favors extended conformations. It is reasonable to expect that the residues in the center of the 10 RU chains will more accurately reflect the behaviors of the corresponding polymers, and there may be environmental conditions (e.g. increasing ionic strength) that favor more order in the polysaccharide chains. Fig. 3.5 illustrates the consequences when the dominant linkage conformation is extrapolated to an 18 RU polysaccharide. For MenA, a repeated L1 linkage conformation produces an extended slightly twisted strand (Fig. 3.5a) with the

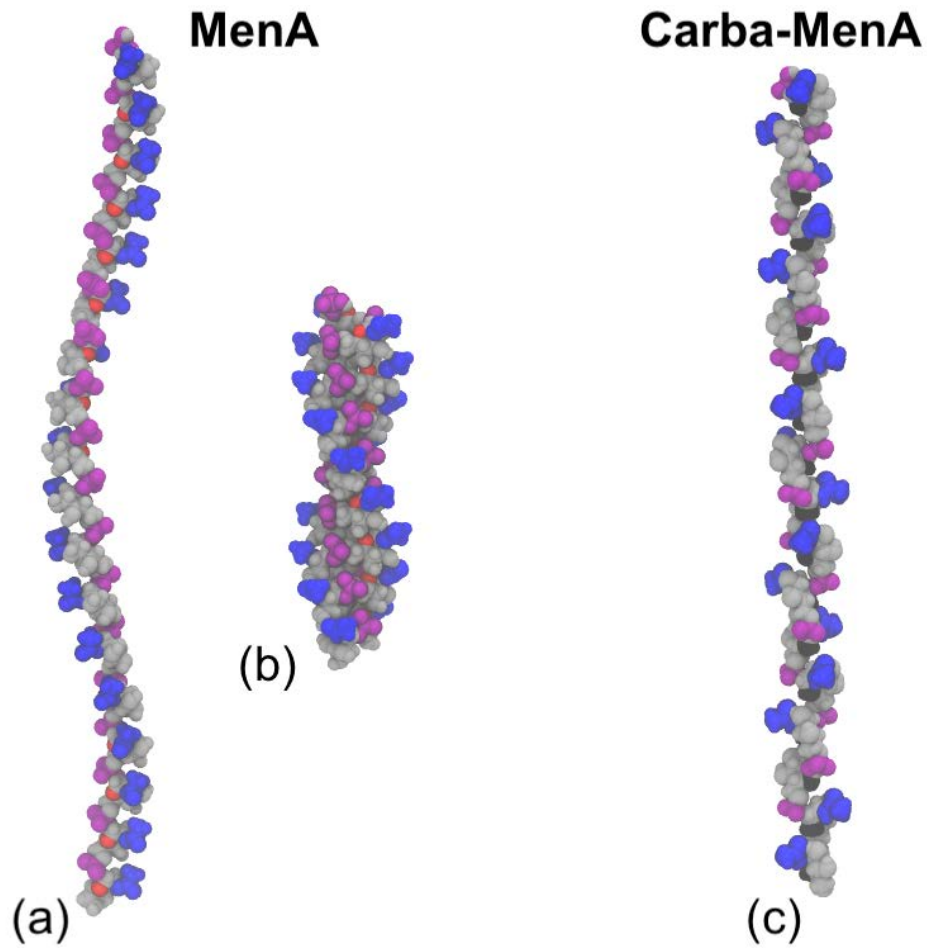


Figure 3.5: Extrapolation of the dominant linkage orientations to 18 RU polysaccharide conformations in MenA and Carba-MenA. N-acetyl groups are shown in blue and phosphate in purple. The ring O5 in MenA is shown in red, whereas the ring methyl in Carba-MenA is indicated in black. The L1 conformation when extrapolated to 18 RU produces the extended twisted strand shown in (a), whereas the hairpin-bend L2 conformation produces the short, broad twisted sheet with S-bends shown in (b). The 18 RU conformation corresponding to cL1 for Carba-MenA, also produces an extended conformation (c), but with a different arrangement of the N-acetyl (blue) and phosphate (purple) groups.

N-acetyl and phosphate groups arrayed on opposite edges. In contrast, the tight hairpin bend L2 orientation of the glycosidic linkage in MenA produces the S-bend conformation when repeated in a polysaccharide, which results in a broad, compact helix with a dramatically shorter end-to-end distance. As a comparison, Carba-MenA not only lacks this chain conformation, but the extended dominant polysaccharide structure (Fig. 3.5b) has a different arrangement of N-acetyl and phosphate groups than occurs in the extended structure of MenA.

3.4.3 Chain length effects

For a vaccine, a short fragment (e.g. 2 RU) of MenA CPS must be representative of the full CPS to provide protection against the pathogen. The populations of the ϕ, ψ, and ω torsion angle averages in our 10 RU simulations are largely in agreement with the low-energy conformations of phosphomethyl-monosaccharides reported from quantum mechanical calculations[92] and MD simulations with the GLYCAM06 forcefield.[93] However, the glycosidic linkages in our polysaccharide simulations do not explore all of the energy minima previously predicted for the monomers. We therefore compare the conformations of the central glycosidic linkage in our 10 RU simulations to the conformations of corresponding 2 RU fragments, which reveal both altered populations of the primary conformations and additional conformations in the disaccharides (Fig. 3.6). For MenA, the large population of extended conformations in the 10 RU strand (Fig. 3.6a, L1 40%) is considerably reduced in the disaccharide (Fig. 3.6c, D2 24%) and the hairpin bend conformations considerably increased (Fig. 3.6c, D1 49%). The dimer also has conformations that are not significantly populated in the 10 RU strand (Fig. 3.6c, D3 and D4). The chain length effect is even more pronounced in

Carba-MenA (compare Fig. 3.6b and Fig. 3.6d) which shows little correspondence between the dominant conformations of the 10 RU strand and the Carba-MenA disaccharide. Further, rotations of the ψ, ω, and χ torsion angles occur much more frequently in the disaccharides than in the longer chains. These differences suggest that, for MenA and Carba-MenA, the disaccharide is not an appropriate model of the CPS. This length dependence and the conformational differences between the MenA and Carba-MenA linkages provides a rationale for the limited cross-protection reported from short Carba-MenA fragments against native MenA CPS.[58]

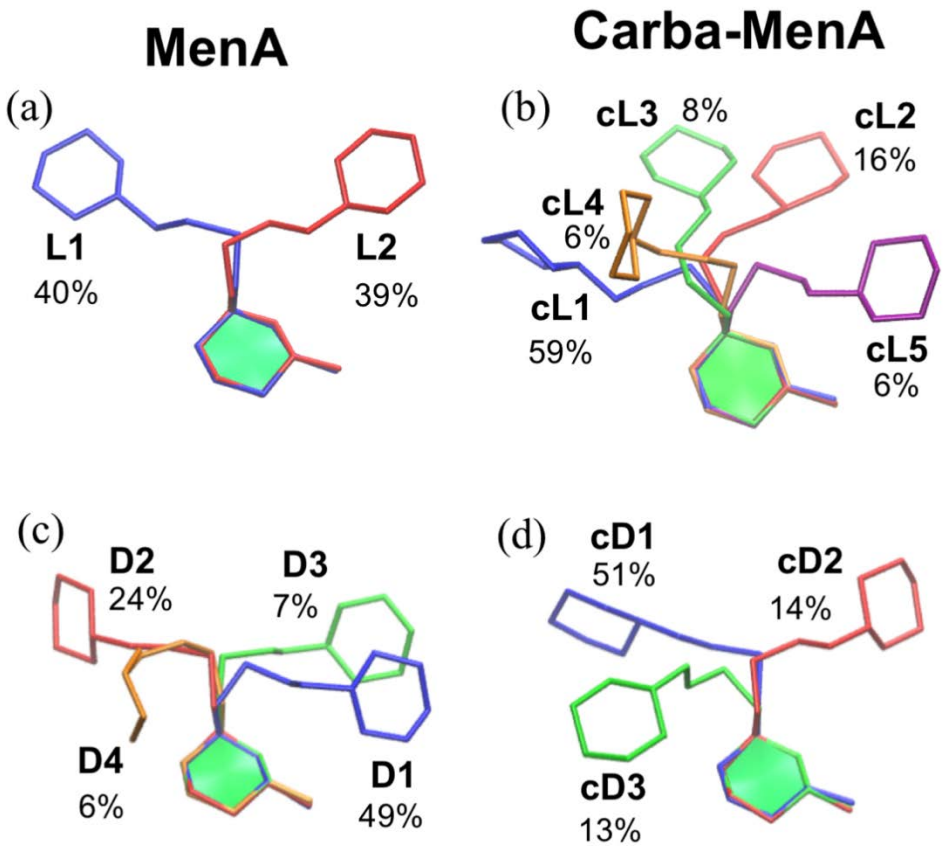


Figure 3.6: Comparison of the dominant conformational families for the central linkage in 10 RU simulations (top row, duplicated from Figure 3.4 for ease of comparison) and the simulations of the corresponding dimers (bottom row) of MenA (left column) and Carba-MenA (right column).

3.5 Conclusions

Despite their chemical similarity, our simulations indicate that substitution of the ring oxygen for a methylene group causes significant changes in the conformations of Carba-MenA as compared to MenA. This is likely due to distinct differences in local electron density: the two lone pairs of electrons on the ring oxygen in the MenA RU are replaced by two hydrogen atoms on an aliphatic carbon for Carba-MenA, resulting in significantly different stereoelectronic constraints on the conformations of the adjacent glycosidic linkage. Alterations in the orientations of the flexible α(1->6) phosphodiester linkage for Carba-MenA as compared to MenA are manifest as distinct differences in the dominant conformations of the central 4 RU, as well as altered conformational distributions of the 10 RU strand. Overall, MenA is more conformationally defined than Carba-MenA, with oscillations between S-bend and extended orientations of the strand. The compact S-bend conformation is absent in Carba-MenA, which displays predominantly extended conformations. Further, our simulations suggest that the disaccharides are not good models for the corresponding 10 RU chains. While a single disaccharide can give a good indication of the range of possible orientations of a glycosidic linkage, it does not necessarily correctly predict their population distribution in the polysaccharide, so care must be taken when extrapolating to longer chains. However, although we found neither saccharide to form stable helices in solution as predicted by the phosphomethyl-monosaccharide simulation, there may be environmental conditions (such as increased ionic strength) that favor stable helices – a topic for future investigations.

3.6 Experimental

The torsion angles describing the conformation of the α(1->6) phosphodiester linkage of native MenA and Carba-MenA are here defined as: ϕ = H1-C1-O1-P; ψ = C1-O1-P-O6'; ω = O1-P-O6'-C6'; ε = P-O6'-C6'-C5' and χ = O6'-C6'-C5'-O5' (MenA) or χ = O6'-C6'-C5'-C7' (Carba-MenA).

3.6.1 Molecular dynamics

We extended our previous simulations of 2 RU and 10 RU of the MenA CPS[1] to 1.5 μs and 1 μs, respectively and ran the same simulations for Carba-MenA. Simulations were performed with version 2.12 of the NAMD molecular dynamics program,[33] using CUDA extensions for calculation of long-range electrostatic forces and non-bonded forces on graphics processing units.[69] The glycans were modeled with the CHARMM36 additive force field for carbohydrates,[31, 32] incorporating *ad hoc* extensions to represent the phosphodiester linkages. Water was simulated with the TIP3P water model,[70] which is compatible with the CHARMM36 carbohydrate force field.

The initial structures for all simulations were built using our CarbBuilder software.[71, 72] The starting conformations for the 10 RU simulations were set to the most common torsion angle values obtained from the 2 RU simulations. All initial structures were subjected to 10,000 steps of standard NAMD minimization in vacuum, followed by solvation in a cubic water box using the Visual Molecular Dynamics (VMD) software.[73] The 2 RU simulations used a 40 x 40 x 40 Å cubic cell (2,000 water molecules) and the 10 RU periodic cubic cells had a side length of 80 Å (16,000 water molecules). Sodium

ions were randomly distributed in each simulation using VMD, to neutralize the negative charge from each phosphodiester.

All MD simulations began with a 122,000 step minimization-and-heating phase which consisted of 5 K temperature reassignments from an initial 10 K up to a temperature of 310 K, with 1,000 steps of minimization and 1,000 steps of MD at each temperature.

The MD equations of motion were integrated with a Leap-Frog Verlet integrator with a step size of 1 fs and periodic boundary conditions were employed. Simulations were run under the isothermal-isobaric (nPT) ensemble at 310 K, which was sustained by a Langevin piston barostat[33] and a Nose-Hoover thermostat (a combination of the Nose-Hoover constant pressure method[74] with piston fluctuation control implemented using Langevin dynamics[75] as implemented in NAMD). Long range electrostatics were calculated with particle mesh Ewald (PME)[76] summation using k = 0.20 $+^{-1}$ and PME grid dimensions were set to be equal to the periodic cell dimensions of the system. Non-bonded interactions were truncated at 15.0 Å with a switching function employed between 12.0 and 15.0 Å. The 1-4 interactions were not scaled, in accordance with CHARMM force field recommendations.

3.6.2 Data analysis

For each simulation, molecular conformations were extracted at 25 ps intervals. The first 200 ns of each simulation was discarded as equilibration, to allow for the flexible polysaccharides to relax from their initial conformations. This left production runs of 1,300 ns for the 2 RU molecules and 800 ns for the 10 RU molecules for conformational analysis.

End-to-end distances were measured as: *r1* being the distance from C1 of the reducing end to C5, and *r2* the distance from C1 of residue 7 to C5 of residue 4. Distances and torsion angles were extracted using VMD's Tcl scripting interface. Statistical values were calculated with in-house Python scripts and graphs. Molecular conformations extracted from the MD simulations were depicted with VMD, where necessary using the PaperChain visualization algorithm for carbohydrates[77] to highlight the hexose rings.

Clustering analysis was performed with VMD's *cluster* command, which employs the quality threshold algorithm.[78] For all clustering calculations, the first 200 ns of each simulation was excluded. All conformations were first aligned on the ring atoms of a single residue: the sixth residue in the 10 RU chain, and the second residue in the disaccharide simulation. Clusters were then calculated with a RMSD fit to the non-hydrogen atoms in the selection (the central 2 RU or 4 RU for the 10 RU strands, the entire 2 RU for the disaccharides). The cut-off value for clusters was set to 3.5 Å for all 2 RU clustering calculations and 5 Å for all 4 RU clustering calculations. Small clusters comprising less than 5% of simulation frames were ignored in the analysis.

Acknowledgements

Computation was performed using the facilities provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za. This work is based on research supported in part by the National Research Foundation of South Africa (Grant Numbers: 111704, 109643). JH thanks the University of Cape Town and the National Research Foundation for scholarship funding.

Chapter 4

Conformations of *S. flexneri* serogroup 2: optimizing the design of a serotype 2a vaccine

4.1 Foreword

The work in this chapter and the next addresses the *Shigella flexneri* bacterium - the dominant pathogen causing diarrheal disease in Africa, particularly in children. Current vaccine development is focused on the serotype-defining glycan RUs of the cell-surface lipopolysaccharides (LPSs) and these are the polysaccharide chains modeled in this work. As no vaccine is currently licensed to prevent disease, initial efforts are specifically focused toward developing a vaccine against serogroup 2 – the cause of the majority of shigellosis infections globally.

The following article "*The Effects of Glucosylation and O-Acetylation on the Conformation of* Shigella flexneri *Serogroup 2 O-Antigen Vaccine Targets*" analyzes the effects of O-acetylation and glucosylation on the conformations of the serogroup 2 O-antigens relative to the unsubstituted backbone (serotype Y). We find important conformational differences between serotype 2a and 2b that suggest only partial cross-protection will occur. However, the O-acetylated serotype 2a is discovered to

have intermediate conformations between 2a and 2b and, therefore, could be leveraged as an optimized antigen with broad coverage against serogroup 2 infection.

The Effects of Glucosylation and O-Acetylation on the Conformation of Shigella Flexneri *Serogroup 2 O-Antigen Vaccine Targets*

Jason Hlozek[a], Neil Ravenscroft[a] and Michelle M. Kuttel[b]*

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] Department of Computer Science, University of Cape Town, Rondebosch 7701, South Africa

* Corresponding author: Tel: +27 21 650 5107; E-mail address: mkuttel@cs.uct.ac.za (Michelle M. Kuttel)

Supplementary Data: Supplementary Figure S4.1, Supplementary Figure S4.2, Supplementary Figure S4.3

4.2 Abstract

Shigellosis is an enteric disease with high morbidity and mortality, particularly in developing countries. There is currently no licensed vaccine available. Most infection is caused by *Shigella flexneri*, of which 30 serotypes have been recognized based on O-antigen polysaccharide structure. Almost all *S. flexneri* serotypes share the same repeating unit backbone (serotype Y), with varying glucosylation, O-acetylation and phosphorylation. The O-antigen is the primary vaccine target; the vaccine valency (and hence cost) can be reduced by cross-protection. Our planned systematic conformational study of *S. flexneri* starts here with 2a, the dominant cause of infection globally.

We employ microsecond molecular dynamics simulations to compare the conformation of the unsubstituted serotype Y backbone with the serogroup 2 O-antigens, to investigate the effect of glucosylation and O-acetylation (O-factor 9) on conformation. We find that serotype Y is highly flexible, whereas glucosylation in 2a restricts flexibility and induces C-curve conformations. Further, the glucose side-chains adopt two distinct conformations, corroborated by the antibody-bound crystal structure data. Additional substitution on O-3 of rhamnose A (whether O-acetylation in 2a or glucosylation in 2b) induces helical conformations. Our results suggest that the O-3-acetylated 2a antigen will elicit cross-protection against 2b, as well as other serotypes containing O-factor 9.

4.3 Introduction

Shigella is one of the five most prevalent diarrheal bacterial pathogens globally, with an annual rate of several hundred million cases of shigellosis infection and up to 800,000 deaths, mainly in sub-Saharan Africa and Asia.[95-97] Due to the extent of this global disease burden, combined with rising antibiotic resistance, *Shigella* is a major worldwide health concern and there is a pressing need for the development of a broad coverage *Shigella* vaccine.[98]

Shigella flexneri accounts for approximately 60% of all shigellosis cases. *S. flexneri* is categorized into 7 serogroups and approximately 30 serotypes, which are characterized by the O-antigen (O-Ag) repeating unit structure expressed on the cell-surface lipopolysaccharide.[99, 100] The O-Ag surface glycans are important vaccination targets.[101] With the exception of serotype 6, all *S. flexneri* O-Ags share the same tetrasaccharide backbone repeating unit (RU), serotype Y:→2)-α-L-Rha*p*^{III}-(1→2)-α-L-Rha*p*^{II}-(1→3)-α-L-Rha*p*^{I}-(1→3)-β-D-Glc*p*NAc-(1→ (Fig. 4.1a). The serotypes differ, with varying glucosylation, O-acetylation and phosphorylation giving rise to the serotypes observed.[99, 102, 103] For *Shigella*, the common backbone and shared substitution patterns raise the possibility of developing a vaccine with broad coverage, which would reduce vaccine complexity and costs.[98, 104-106] The Global Enteric Multicenter Study (GEMS) suggested that a quadrivalent vaccine — containing *S. flexneri* serotypes 2a, 3a and 6 as well as the sole serotype of *S. sonnei*— could provide direct protection against 64% of *Shigella* strains causing infection in children in low-income areas, with cross-reactivity potentially extending this to 88% overall coverage.[95, 107]

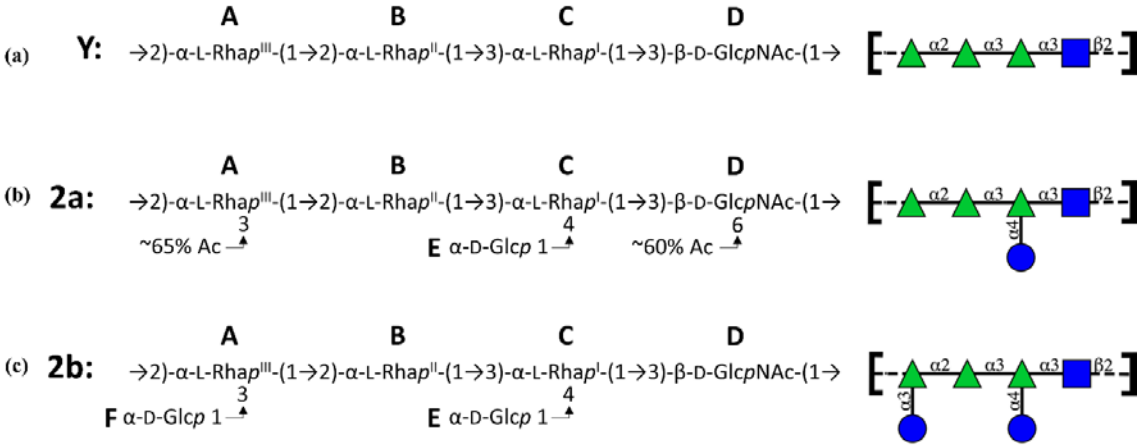


Figure 4.1: Line structures and schematic diagrams of the O-Ag repeating units of Shigella flexneri (a) serotype Y, (b) serotype 2a, and (c) serotype 2b. Serogroup 2 all have the same tetrasaccharide backbone repeating unit as serotype Y, with glucosylation on O-4 of rhamnose C. Serotype 2a has non-stoichiometric O-acetylation in two positions and serotype 2b has glucosylation on rhamnose A. Diagrams are depicted using the ESN symbol set[108] with: green triangle - Rha, blue square - GlcNAc, blue circle - Glc.

We focus here on serotype 2a, which causes the majority of *S. flexneri* infection and is thus the primary target of current vaccine development efforts.[109, 110] The 2a O-Ag adds glucosylation on O-4 of rhamnose C to the common backbone (serotype Y, Fig. 4.1a). The 2a native polysaccharide (Fig. 4.1b) includes partial O-acetylation at O-3 of rhamnose A (O-factor 9, ~65%), O-4 of rhamnose A (~25%) and O-6 of GlcNAc (O-factor 10, ~60%). However, the level and position of O-acetylation reported can be affected by the strain, method of isolation/purification and storage conditions for the O-Ag.[99] Serotype 2b has additional glucosylation at O-3 of rhamnose A and no O-acetylation (Fig. 4.1c).

Similarity in antigen primary structure does not necessarily result in strong cross-reactivity between closely related antigens, as small differences in glycan structure can

significantly alter chain conformation and molecular flexibility. Therefore, conformational analysis can play an important role in investigating cross-reactivity between antigens.[22]

For serogroup 2, studies in the 1990s showed limited cross-protection conferred by 2a against both serotype 2b and serotype Y.[104, 106] However, the immunogens were not yet fully characterized and the 2a O-acetylation pattern was only determined in 2007.[111] Recent clinical trials with non-acetylated 2a conjugate vaccines have raised functional antibodies against 2a, but cross-protection against other serotypes has not been reported.[110, 112]

In general, O-acetylation has been identified as an important epitope in *S. flexneri* 2a cross-reactivity and is in need of further investigation.[113] In particular, O-3 acetylation (designated O-factor 9)[99] significantly affects binding of monoclonal antibodies, suggesting the importance of this epitope for antigenicity, while O-6 acetylation on GlcNAc (O-factor 10)[99] has only a minor effect on antibody binding.[114] Indeed, O-factor 9 is the likely reason for antibody cross-reactivity observed against serotype 6 following 2a conjugate vaccination in humans.[115] O-factor 9 is also present on *Shigella* serotypes $1a_1$, 1b, $5a_1$, Y_1 and $7a_1$,[99, 116] which may offer broader *S. flexneri* coverage.

Molecular dynamics (MD) is increasingly used to investigate glycan antigen conformations in order to compare related serotypes and rationalize cross-protection phenomena.[1, 117-121] Previous MD simulations of 4 RU of the *Shigella* serotype Y polysaccharide indicate a flexible backbone with a diverse array of conformations.[122] Short simulations (60 – 350 ns) of twelve *S. flexneri* O-Ags showed very similar backbone linkage conformations for 3 RU strands across the different serotypes: the

effect of substituents was to limit the conformations that the backbone could adopt.[123] However, chain length may be a factor, as differing conformations for varying chain lengths have been observed in MD simulations of glycan antigens[1, 2] and, for flexible glycans, simulations of over 1 μs may be necessary to replicate native polysaccharide behavior.[120]

There remain important conformational questions for the *S. flexneri* serogroup 2 O-Ags, as follows. Does serotype 2a (glucosylation on rhamnose C) have altered conformation and flexibility relative to the serotype Y (unbranched backbone)? Does O-3 acetylation on rhamnose A (O-factor 9) alter the conformation of a non-acetylated serotype 2a? Does serotype 2b (with additional glucosylation on rhamnose A) have a different conformation to serotype 2a?

To investigate these questions, we compare microsecond MD simulations of 6 RU polysaccharides of the serogroup 2 O-Ags (comprising serotype 2a, 2a with O-3 acetylation on rhamnose A, and 2b) with the unsubstituted backbone O-Ag (serotype Y). The conformations of 2a and Y are corroborated by comparison with published X-ray crystal structures of monoclonal antibodies bound to O-Ag fragments. Comparison of the polysaccharide conformations from these simulations is discussed in light of potential cross-reactivity to inform future vaccine design. This work is the first stage of our planned systematic study of the conformations of key *Shigella flexneri* O-Ags.

4.4 Materials and methods

The glycosidic linkages in the *S. flexneri* polysaccharides are described by two torsion angles defined as: ϕ = H1-C1-O1-C_x' and ψ = C1-O1-C_x'-H_x'. These torsion angles are

analogous to ϕ_H and ψ_H in IUPAC convention and are consistent with our previous modeling studies.[2, 27, 124] Our established systematic approach to modeling of polysaccharide antigens involves first determining the preferred conformations of each glycosidic linkage in the polysaccharide by calculation of the ϕ, ψ potential of mean force (PMF) for the corresponding disaccharide linkages and then progress to molecular dynamics simulations of oligosaccharides in aqueous solution to establish the preferred conformations and dynamics of the glycan chains.[1, 65, 120]

4.4.1 Disaccharide PMF calculations

We identified the preferred conformations of the each of the glycosidic linkages by calculation of the potential of mean force (PMF) for rotation about the ϕ and ψ torsion angles. PMFs were calculated using the metadynamics[36] routine incorporated into NAMD[33], with the glycosidic linkage torsion angles used as collective variables. All PMF surfaces were calculated in the gas-phase.

4.4.2 Molecular dynamics

All simulations were performed using the NAMD software package[33] with CUDA extensions for accelerated calculation of long-range electrostatic potentials and non-bonded forces on graphics processing units.[69] All structures were built with our in-house CarbBuilder software[71, 72] and modeled with the CHARMM36 additive force field for carbohydrates.[31, 32] The TIP3P model[70] was used to simulate water.

Initial chains of 3 RU of the *S. flexneri* O-Ags (not discussed here) were built with torsion angles set to the respective energy minima calculated in the PMFs. The 3 RU simulations were run in solution for 300 ns and the most common torsion angles from

these simulations were used to build the longer 6 RU polysaccharides reported in this study. All starting structures underwent 10,000 steps of standard NAMD minimization in vacuum and were then solvated in a cubic water box with the *solvate* command in the Visual Molecular Dynamics (VMD) program.[73] The 6 RU structures were placed into a cubic water box with side lengths of 90 Å (~23,000 water molecules). Periodic boundary conditions were employed for the solvated simulations. All solvated MD simulations first underwent a minimization-and-heating protocol of 122,000 steps, consisting of 5 K incremental temperature reassignments beginning at 10 K up to 310 K, with 1,000 steps of NAMD minimization and 1,000 steps of MD at each temperature reassignment.

A Leap-Frog Verlet integrator was used to integrate the MD equations of motion using a step size of 1 fs. Simulations were sampled under the isothermal-isobaric (nPT) ensemble at 310 K, which was achieved with the use of a Langevin piston barostat[33] and a Nose-Hoover thermostat (a hybrid of the Nose-Hoover constant pressure method[125] with piston fluctuation controlled using Langevin dynamics[75] as implemented in NAMD. Long-range electrostatics were implemented with particle mesh Ewald (PME) summation[76] with k = 0.20 $Å^{-1}$ and PME grid dimensions that were set to 90 Å for the 6 RU simulations. Non-bonded forces were truncated at 15.0 Å with a switching function applied from 12.0 Å to 15.0 Å. The 1-4 interactions were not scaled, in accordance with CHARMM force field recommendations.

4.4.3 Data analysis

Molecular conformations were sampled at 12.5 ps intervals. The simulations equilibrated for 200 ns, after which production runs of 800 ns were performed. Inter-

atomic distances and torsion angles were extracted using VMD's Tcl scripting and graphical user interfaces. The end-to-end distance, *r*, was measured between C2 of rhamnose B at the non-reducing end and C1 of rhamnose C at the reducing end, excluding the more flexible terminal residues which are less representative of the native O-Ag polymer. Statistical measures were calculated with in-house Python scripts. For the antibody-bound serotype 2a conformations, the median torsion angle of the four decasaccharides was plotted on the respective ϕ/ψ plots. Heat maps were generated from measurements of the ϕ/ψ torsion angles from the glycosidic linkages of both central repeating units, where the torsion angles from each frame were sorted into bins with a size of 1 degree in each dimension.

VMD[73] was used to visualize the extracted molecular conformations, making use of the PaperChain visualization algorithm[77] to highlight the carbohydrate rings. Conformational clusters were calculated from the 800 ns production runs with VMD's internal *cluster* command. Before clustering, the time series trajectory snapshots were first aligned on the ring atoms of the 'B-C' disaccharide fragment of RU3 – a central fragment with the least linkage flexibility. Clusters were then calculated with a RMSD fit to the non-hydrogen atoms of the central 4 RUs with the cut-off criterion set to 6.5 Å.

4.5 Results

We begin with a broad comparison of the chain extension and flexibility of the *S. flexneri* backbone (serotype Y) with the serogroup 2 O-Ags. We then explore the effects of glucose side-chains and rhamnose O-acetylation on the backbone

conformation. Finally, we link the differences in serogroup 2 conformations to altered orientations of specific backbone glycosidic linkages.

To investigate the issue of convergence in the simulations, we performed block averaging analysis[35, 126] of the end-to-end distance, *r*, and the radius of gyration, R_g, for the 6 RU chains (Supp. Fig. S4.1a and S4.1b). Overall, the blocked standard errors reach a plateau for both measures in each O-Ag simulation, indicating convergence.[50] Correlation times of less than 45 ns for all serotypes indicate that 200 ns is a sufficient equilibration time and also that the 6 RU O-Ags are adequately sampled over the course of the 800 ns production runs.

4.5.1 O-Ag chain extension and flexibility

The end-to-end distance, *r*, is a measure of glycan chain extension and the *r* time series (a plot of *r* versus time) is a simple visualization of molecular flexibility. We define *r* for all O-Ags as the distance between C2 of rhamnose B in RU1 and C1 of rhamnose C in RU6 for the glycan backbone (Fig. 4.2a).

The *r* time series for serotype Y O-Ag (Fig. 4.2b) demonstrates the extreme flexibility of the unbranched backbone, which rapidly transitions through a broad range of values from 7 Å to 74 Å. The corresponding *r* histogram (Fig. 4.2c) has two peaks: a population of extended conformations (peak at *r* = 54 Å) comprising 59% of the 800 ns simulation and a population of compact conformations (peak at *r* = 24 Å) comprising 41% of the simulation.

In contrast, the *r* time series for the 2a O-Ag (Fig. 4.2d) has a more restricted range (5 Å to 57 Å), with smaller values of *r* dominating throughout the simulation. The

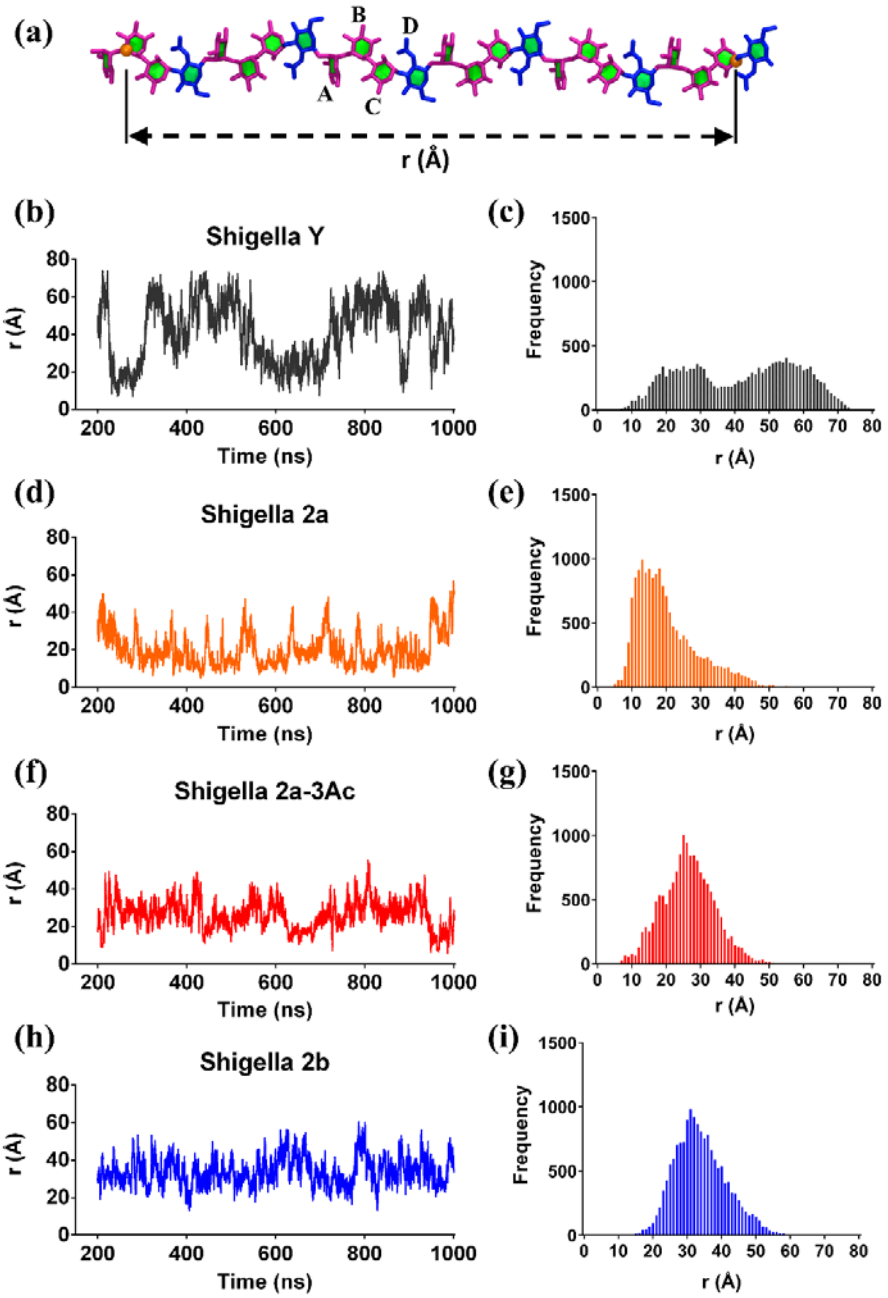


Figure 4.2: Comparison of the r time series and distribution for 6 RU of serotype Y and serogroup 2 O-antigens, excluding the first 200 ns of each simulation. (a) A model of the 6 RU Y chain with the end-to-end distance, r, indicated; rhamnose is colored pink and N-acetyl-glucosamine blue. The line graphs show the r time series (left column) and the bar graphs the corresponding histograms (right column) for Y (grey, b and c); 2a (orange, d and e); 2a with O-3 acetylation (red, f and g); and 2b (blue, h and i).

corresponding histogram for serotype 2a (Fig. 4.2e) shows a unimodal, right-skewed distribution with a peak at $r \approx 14$ Å (mean 20 Å, $\sigma = 9$ Å). This difference between the Y backbone and the branched 2a O-Ag indicates that glucosylation of the backbone has a conformational effect, resulting in a predominance of compact, curved conformations of the backbone in 2a.

The 2a-3Ac O-Ag has similar flexibility to 2a, with values of *r* ranging from 6 Å to 55 Å (Fig. 4.2f), but O-acetylation at position 3 of rhamnose A (O-factor 9) shifts *r* to a symmetrical unimodal distribution with a peak at $r \approx 26$ Å, $\sigma = 8$ Å (Fig. 4.2g). Thus, O-factor 9 has direct conformational consequences, resulting in more extended conformations for the O-acetylated 2a O-Ag.

The 2b O-Ag is similarly flexible, with *r* ranging from 13 Å to 60 Å (Fig. 4.2h). The additional glucosylation at position 3 of rhamnose A also shifts the values of *r* relative to the 2a chain: the *r* distribution has a peak at $r \approx 33$ Å, $\sigma = 8$ Å (Fig. 4.2i). Thus, glucosylation at O3 has an equivalent effect on chain extension to O-3 acetylation: compare the similar profiles for the corresponding histograms Fig. 4.2g and 4.2i. The reasons for these fairly dramatic shifts in chain extension are revealed in a detailed examination of the chain conformations for the O-Ags, as discussed below.

4.5.2 O-Ag chain conformations

The dominant conformational families and associated frequencies for each O-Ag chain are shown in Fig. 4.3. Here we consider only the conformation of the middle 4 RU of each O-Ag, as the ends of the 6 RU chain are less representative of the conformation of a *S. flexneri* O-Ag polysaccharide.

The extreme flexibility of the serotype Y antigen (unbranched backbone) is demonstrated by the high number of conformational families observed for this O-Ag (Fig. 4.3a). The two peaks in the *r* histogram for the Y O-Ag (Fig. 4.2c) are not associated with specific conformations, but rather reflect a range of both compact and extended conformations. As expected from the *r* histogram, the predominant cluster (Y1) is an extended conformation (*r* = 54 Å). This is in agreement with an early conformational prediction by the HSEA method.[127] However, cluster Y1 encompasses just 25% of the simulation conformations, with a smaller distinct extended family (Y3, with a tight bend in the middle of the chain) representing a further 12% of the simulation's conformations. Similarly, there are a range of compact conformations (Y2, Y4) that contribute to the peak at *r* = 24 Å in the histogram. This diverse range of conformations, together with their low occupancies, indicates that the Y O-Ag is a highly flexible polysaccharide. Our results are in good agreement with a previous MD study of serotype Y, also finding a wide array of conformations in shorter chains of up to four repeating units.[122]

For the 2a O-Ag, the *r* time series indicates a flexible molecule, but with a shift to more compact conformations relative to the Y backbone. Indeed, the dominant conformation (Fig. 4.3b, cluster 2a-1) is a compact C-curve with ends in close proximity, representing 38% of the simulation. The second conformational cluster is a similar, but slightly distorted, curve (2a-2, 17%). The final small cluster (2a-3) is a helical conformation, but occupies only 13% of simulation time. There are no significant families of extended conformations for 2a, as expected from the skewed distribution for the *r* histogram.

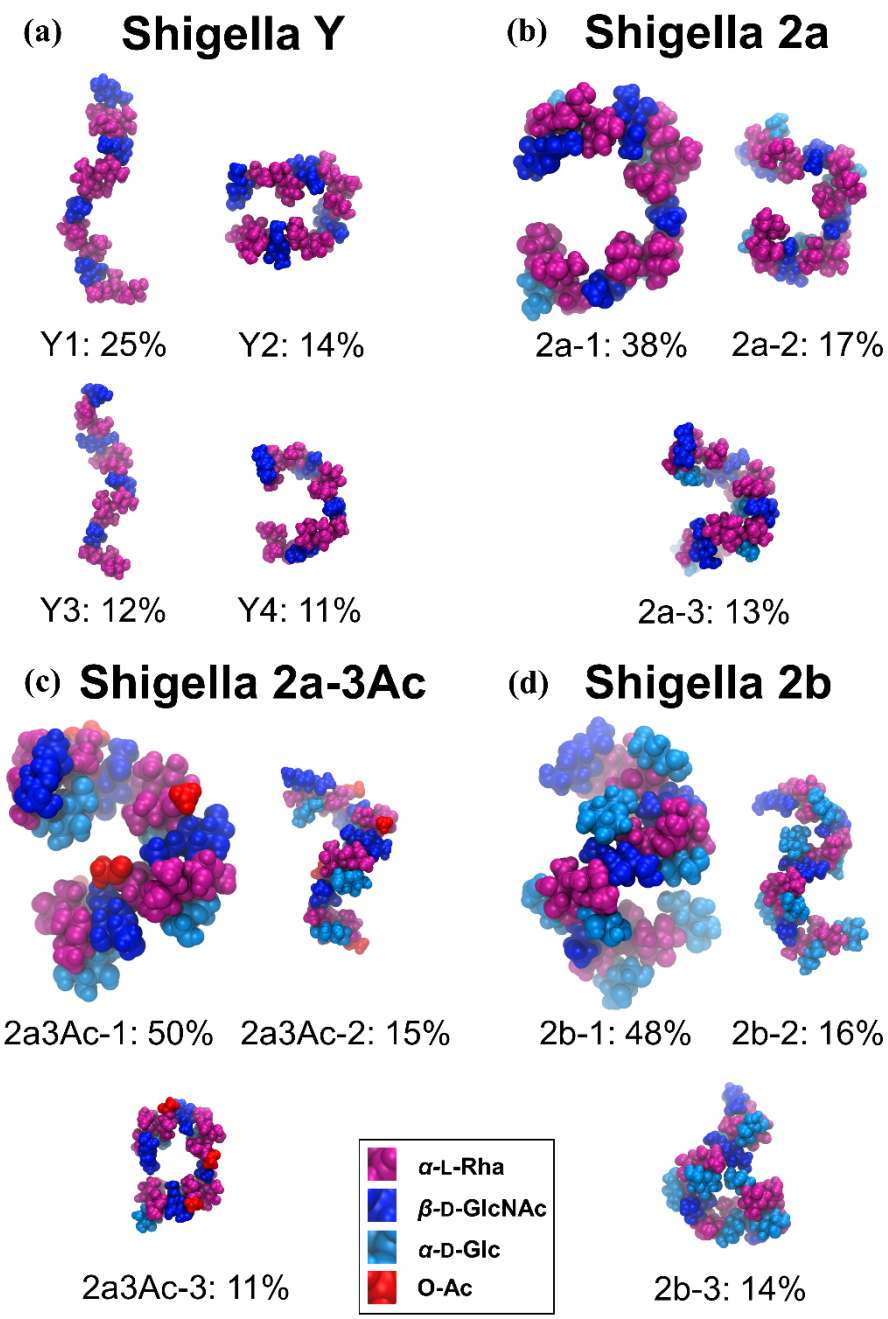


Figure 4.3: Conformational families of the central 4 RU of the 6 RU chains: (a) Y, (b) 2a, (c) rhamnose A O-3 acetylated 2a and (d) 2b. The relative percentage occupancy of each conformational family during the 800 ns production run (ignoring the initial 200 ns) is indicated; clusters below 7% are not shown. The sugars are colored as indicated in the legend: pink for Rha, dark-blue for GlcNAc, cyan for Glc side chains, and red for O-acetyl groups.

Conformational analysis reveals that O-3-acetylation on rhamnose A (O-factor 9) in the 2a O-Ag causes a significant conformational change in the 2a backbone from a flexible C-curve to a more stable helix (Fig. 4.3c), resulting in the more extended conformations seen in the *r* histogram (Fig. 4.2g). The helix of the primary conformational cluster has 3 RU per turn and a helical pitch of 18 Å, comprising 50% of the simulation time (Fig. 4.3c, 2a3Ac-1). A secondary cluster occupies 15% of the simulation time and represents a more extended helix, also with 3 RU per turn but with a pitch of 28 Å (Fig. 4.3c, 2a3Ac-2). A curved conformation similar to the de-O-acetylated 2a O-Ag occurs for only 11% of the simulation (Fig. 4.3c, 2a3Ac-3). The O-acetyl groups are also exposed on the outside edge of the helices, and so would be accessible for antibody binding.

Glucosylation at position 3 of rhamnose A has equivalent conformational consequences to O-3-acetylation, favoring more stable helical conformations over looser curved conformations (Fig. 4.3d). Indeed, comparison of the two primary conformational clusters of the 2b O-Ag (Fig. 4.3d) with the 2a-3Ac O-Ag (Fig. 4.3c) shows similar helical conformations and populations over the 800 ns simulation. The chief 2b conformational cluster, 2b-1, is comparable to 2a3Ac-1: a 3 RU-per-turn helix with a pitch of 18 Å. The minor clusters are representative of extensions (cluster 2b-2) and compressions (2b-3) of this helix. Therefore, these simulations show that substitution at position 3 of rhamnose A, be it O-acetylation or glucosylation, results in a helical conformation of the serogroup 2 O-Ags.

4.5.3 O-Ag glycosidic linkage conformations

Conformational changes in an O-Ag chain are brought about primarily from changes in the orientations of the constituent glycosidic linkages, because the ring conformations of the monosaccharides are typically relatively constrained. The range of motion and the dominant orientations of each glycosidic linkage in the central RU of each O-Ag strand over the course of the simulations is depicted with heat maps of the linkage torsion angles, ϕ and ψ, shown in Fig. 4.4.

All the backbone glycosidic linkages – in the Y, 2a and 2b O-Ags – show a narrow range of ϕ torsion values centered on ϕ ≈ 40° and more flexibility in the ψ torsion angles, with two primary ψ conformations at ψ ≈ 14° and ψ ≈ -35° (from here on referred to as +ψ and -ψ). Broadly, the torsion angle populations for the unbranched Y O-Ag (Fig. 4.4a) are consistent with previous MD simulations[122] as well as with the published X-ray crystal structure of a monoclonal antibody bound to a Y pentasaccharide.[128] Further, for the non-acetylated 2a and serotype 2b, our linkage analysis is in good agreement with previous simulations of 3 RU chains.[123]

The α-L-Rha*p*^{II}-(1→3)-α-L-Rha*p*^{I} (B-C) linkage is the most constrained of the backbone linkages and is relatively consistent across the serogroup 2 O-Ags (Fig. 4.4, second column). This linkage is moderately affected by steric interactions with the glucosylation at O-4 on residue C, shifting from ϕ, ψ ≈ 40°, 8°(+ψ) in unbranched Y to ϕ, ψ ≈ 40°, -22°(-ψ) in the glucosylated serogroup 2 O-Ags. The neighboring α-L-Rha*p*^{III}-(1→2)-α-L-Rha*p*^{II} (A-B) linkage is more dramatically affected by glucosylation at rhamnose C: this linkage favors ϕ, ψ≈48°, 17° (+ψ) in Y, which shifts to ϕ, ψ ≈39°, -39° (-ψ) in the remaining O-Ags (Fig. 4.4, first column). The combined shift of both the A-B

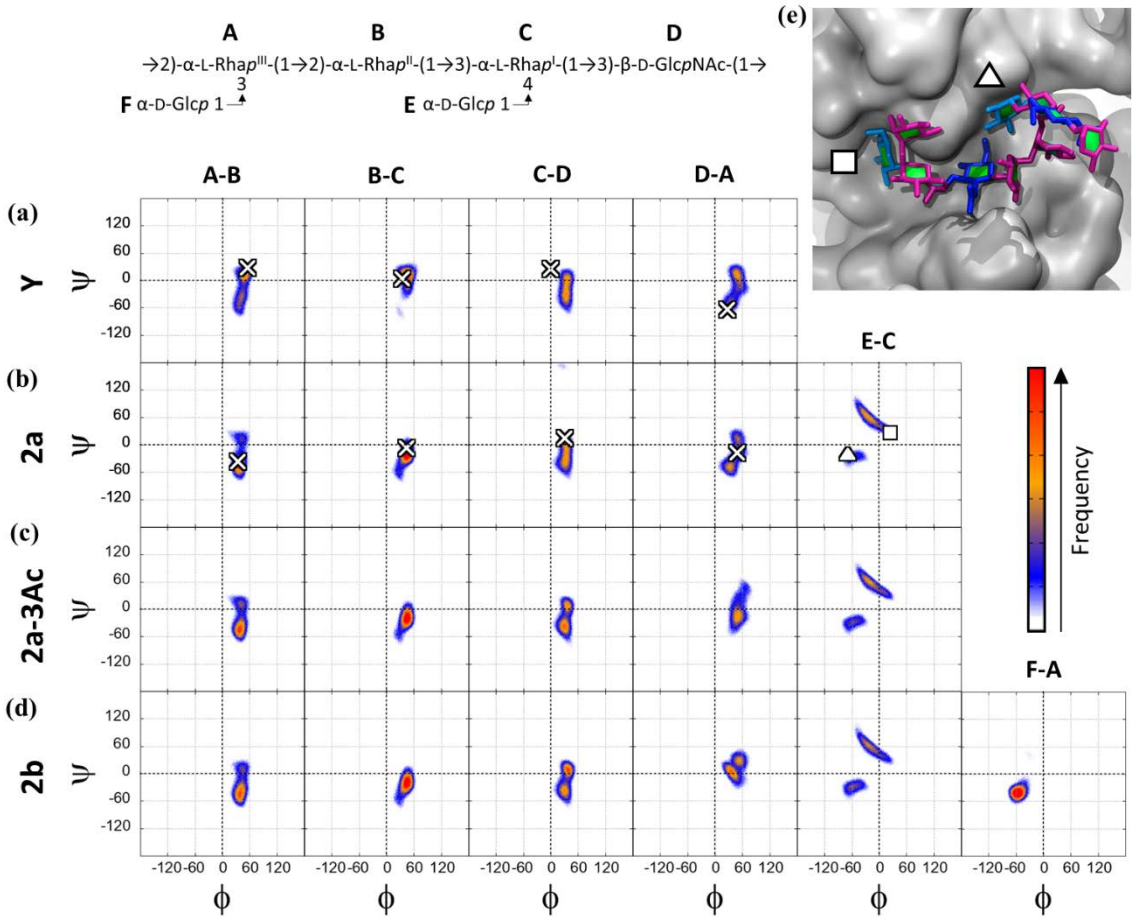


Figure 4.4: Scatter plots illustrated as heat maps, with heat scale on the right, for the torsion angles (φ, ψ) of each glycosidic linkage in the repeating unit: (a) Y, (b) 2a, (c) rhamnose A O-3 acetylated 2a, and (d) 2b. The white cross symbols on the plots of Y and 2a correspond to the torsion angles adopted in the bound states to respective serospecific antibodies in published crystal structures.[128, 129] The E-C linkage of the serogroup 2 O-Ag can adopt two distinct conformations (as indicated for 2a, fifth column). These correspond to a perpendicular (white square) and a face-on (white triangle) orientation.

and B-C linkages to -ψ conformations in the branched O-Ag causes significant curvature in the saccharide chain: a Y backbone with all +ψ linkages is an extended chain (Supp. Fig. S4.2a), whereas all -ψ rotations result in a curved conformation (Supp. Fig. S4.2b). Overall, this shift in glycosidic linkage orientation results in the curved conformations that dominate for glucosylated 2a and 2b O-Ags.

The α-L-Rha*p*^{I}-(1→3)-β-D-Glc*p*NAc linkage (C-D, third column in Fig. 4.4) is the most flexible linkage and its conformation is similar across all O-Ags, with ψ ranging from +30° to -60°. This linkage is therefore the primary source of conformational flexibility, particularly in the branched O-Ags.

The β-D-Glc*p*NAc-(1→2)-α-L-Rha*p*^{III} (D-A, Fig. 4.4, fourth column) linkage shows the greatest difference between the Y, 2a, 2a-3Ac and 2b O-Ags, which is unsurprising as the adjacent α-L-Rha*p*^{I} (A) residue has various substitutions across the four O-Ags. Alterations in the conformation of this linkage are therefore the source of most of the conformational variation between the serogroup 2 O-Ags. The D-A linkage in Y (Fig. 4.4a, fourth column) and 2a (Fig. 4.4b, fourth column) have a similar distribution, with two primary conformations (+ψ and -ψ), indicating a flexible linkage. However, both O-3-acetylation on rhamnose A (2a-3Ac, Fig. 4.4c, fourth column) and 3-glucosylation at rhamnose A (2b, Fig. 4.4d, fourth column) narrow the range of this linkage to a single dominant conformation, with ψ ≈ -20° (2a-3Ac) and ψ ≈ -1° (2b). This is due to interactions between the substituent (O-Ac or glucose) and the N-acetyl group on D, which restricts rotation of the D-A linkage. This additional constraint on the backbone orientation produces the dominant helical conformational families seen for the 2a-3Ac and 2b O-Ags (Supp. Fig. S4.2c). However, rotations in the C-D linkages still allow the helical conformations significant flexibility.

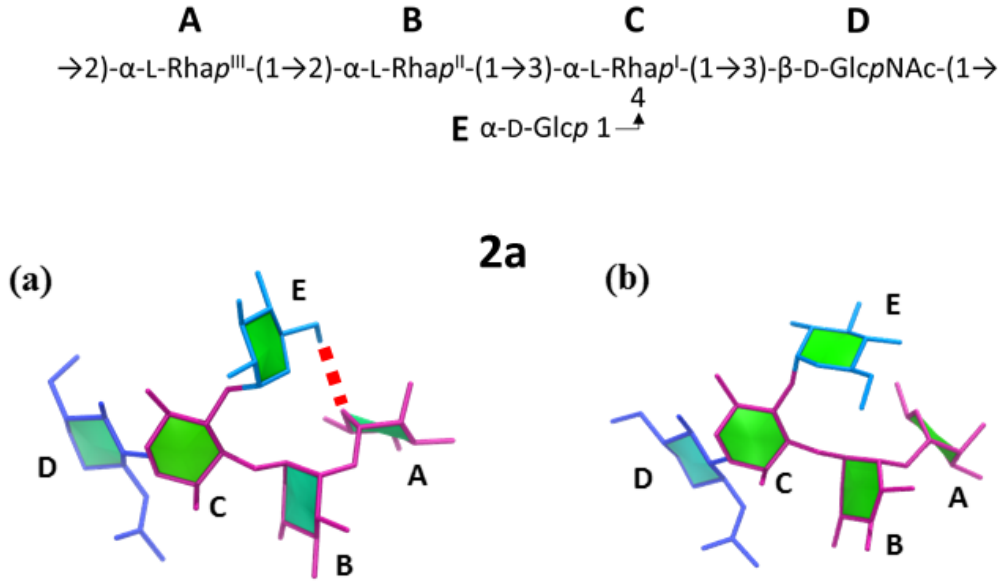


Figure 4.5: The two conformations of the glucose side chain (residue E) on rhamnose C of the serogroup 2 O-antigens: (a) perpendicular orientation (+ψ; white square on bound crystal structure - Fig. 4.4e) and (b) face-on orientation (-ψ, white triangle on bound crystal structure – Fig. 4.4e). The dominant perpendicular orientation is stabilized by a hydrogen bond between the O6 hydroxyl of glucose E and the ring oxygen of rhamnose A, as indicated by the red, dashed line.

The heatmaps for the α-D-Glc*p*-(1→ 3)-α-L-Rha*p*^{I} linkage (E-C, Fig. 4.4, fifth column) are rather unusual, with two disjoint populations: a primary population at ϕ, ψ = -17°, 59° (+ψ, 74%) and a secondary population at ϕ, ψ = -53°, -30° (-ψ, 26%). This is consistent across the 2a, 2a-3Ac and 2b O-Ags. Torsion angles for this linkage in a crystal structure of a monoclonal antibody bound to four decasaccharide (2 RU) chains of de-O-acetylated 2a[129] are in agreement with the simulation populations (Fig. 4.4b – white crosses on heatmaps). Interestingly, the two α-glucose side chains in the decasaccharide (Fig. 4.4e) are each bound to the antibody with a different conformation, corresponding to the two dominant conformations from our MD simulations, illustrated in Fig. 4.5a and 4.5b, respectively. The +ψ state orientation is

partially stabilized by a hydrogen bond between the O-6 hydroxyl of α-glucose (residue E) to the ring oxygen of rhamnose A (Fig. 4.5b). Previous MD simulations of serotype 2 also found two populations for this linkage, albeit with altered occupancies.[123] The disjoint populations for this linkage are a result of high-energy steric clashes in the intermediate conformations, between the primary hydroxyl of the glucose side chain and the backbone. This is evidenced by a comparison of calculated ϕ, ψ PMFs for this linkage in the α-D-Glc*p*-(1→ 3)-α-L-Rha*p* disaccharide (Supp. Fig. S4.3a) and a pentasaccharide RU of 2a (Supp. Fig. S4.3b): the pentasaccharide PMF shows two low-energy conformations (the +ψ orientation is 1 kcal.mol^{-1} lower in energy) separated by a 3-4 kcal.mol^{-1} energy barrier.

4.6 Discussion and conclusions

For this computational investigation into the effects of glucosylation and O-acetylation on the conformations of *S. flexneri* serogroup 2 O-antigens, we use the unbranched serotype Y O-Ag as a reference for comparison. We find that our extended simulation of a longer repeating unit of the serotype Y O-Ag agrees with the diverse array of conformations reported in previous MD simulations[122], as well as with the published X-ray crystal structure of a monoclonal antibody bound to a Y pentasaccharide.[128] Although the unbranched serotype Y O-Ag is very flexible, the glucosylation that defines serogroup 2 (O-4 of rhamnose C) restricts the range of motion of the 2a O-Ag, resulting in flexible C-curve conformations as compared to the more extended chains that predominate in serotype Y. However, the prevalence of C-curve conformations in both 2a and the Y backbone may provide a rationale for the partial cross-reactivity elicited from 2a anti-sera against the Y O-Ag in early immunological studies.[104]

Our simulations show two dominant orientations for the glucose branch on O-4 of rhamnose C (serogroup 2), which we attribute to steric clashes between the primary hydroxyl of the glucose side chain and the backbone. The existence of two orientations for this side chain is corroborated by the presence of the same two orientations in the glucose side chains of a 2 RU antibody-bound X-ray crystal structure.[129]

O-3-acetylation of rhamnose A (O-factor 9) in serogroup 2 limits the flexibility of the O-Ag backbone, changing the dominant conformation from a C-curve to a helix with solvent exposed epitopes. This change provides a rationale for the strong influence of O-factor 9 in antibody binding to a range of synthetic 2a oligosaccharides.[114] Further, this may provide insight into the cross-reactivity observed between antibody raised by an O-acetylated 2a vaccine and serotype 6, which shares the O-factor 9 epitope.[115] O-factor 9 is also present on *Shigella* serotypes $1a_1$, 1b, $5a_1$, Y_1, and $7a_1$[99, 116], providing the possibility of cross-protection with O-3-acetylated 2a.

Glucosylation on O-3 of rhamnose A also restricts the O-Ag flexibility, producing quite well-defined helices for serotype 2b as well. Therefore, we find that substitution at position 3 of rhamnose A, be it O-acetylation or glucosylation, results in a similar helical conformational of the serogroup 2 O-Ag. Overall, this similarity in conformation suggests that the 2a O-Ag with O-factor 9 could be a better antigen for eliciting cross-reacting antibodies against 2b, compared to 2a without O-acetylation. We hope that ongoing pre-clinical or clinical studies on *Shigella* vaccines currently under development will support our findings.

Future investigation is needed into the effect, if any, of acetylation on O-6 of GlcNAc (O-factor 10). These results, showing a significant effect of O-acetyl and glucose

substitution patterns on *Shigella* O-Ag conformations, can inform the rational selection of suitable vaccine antigens to achieve the broadest possible cross-protection against shigellosis.

4.7 Associated content

Supporting Information

Blocked standard error calculations for determining the extent of simulation convergence (Supplementary Figure S4.1); Static backbone models built from dominant linkage orientations (Supplementary Figure S4.2); PMF diagrams for the serotype 2a branched glucose linkage (Supplementary Figure S4.3). This material is available free of charge via the Internet at http://pubs.acs.org.

Author Information

Corresponding Author

*E-mail: mkuttel@cs.uct.ac.za. Phone: +27 21 650 5107.

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Notes

The authors declare no competing financial interest.

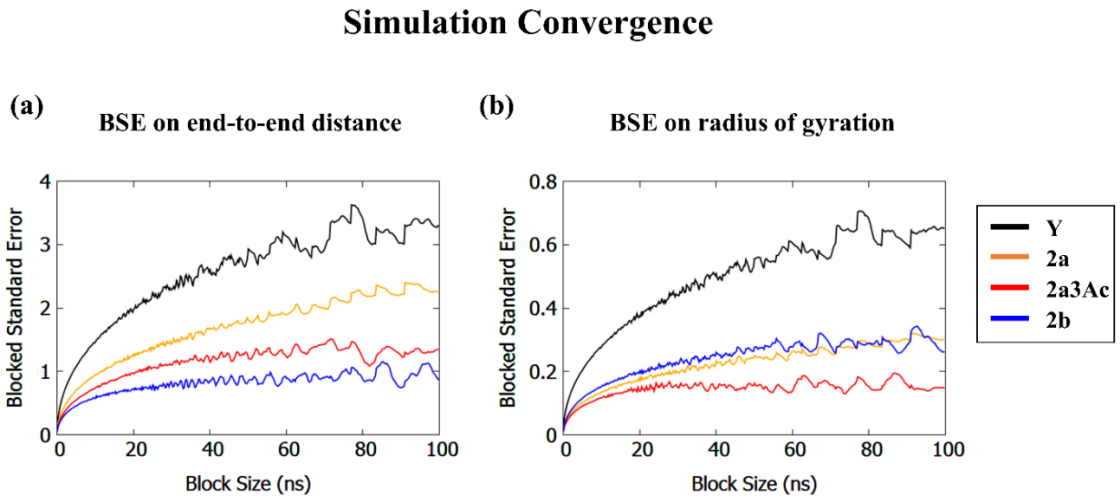
Acknowledgements

Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za. This work was supported by the National Research Foundation of South Africa [111704 to N.R., 103805 to M.K., 109643 to J.H.]; and University of Cape Town PhD scholarship funding. Some of these results were presented as an oral communication at the 25th International Symposium on Glycoconjugates in Milan, Italy (August 2019).

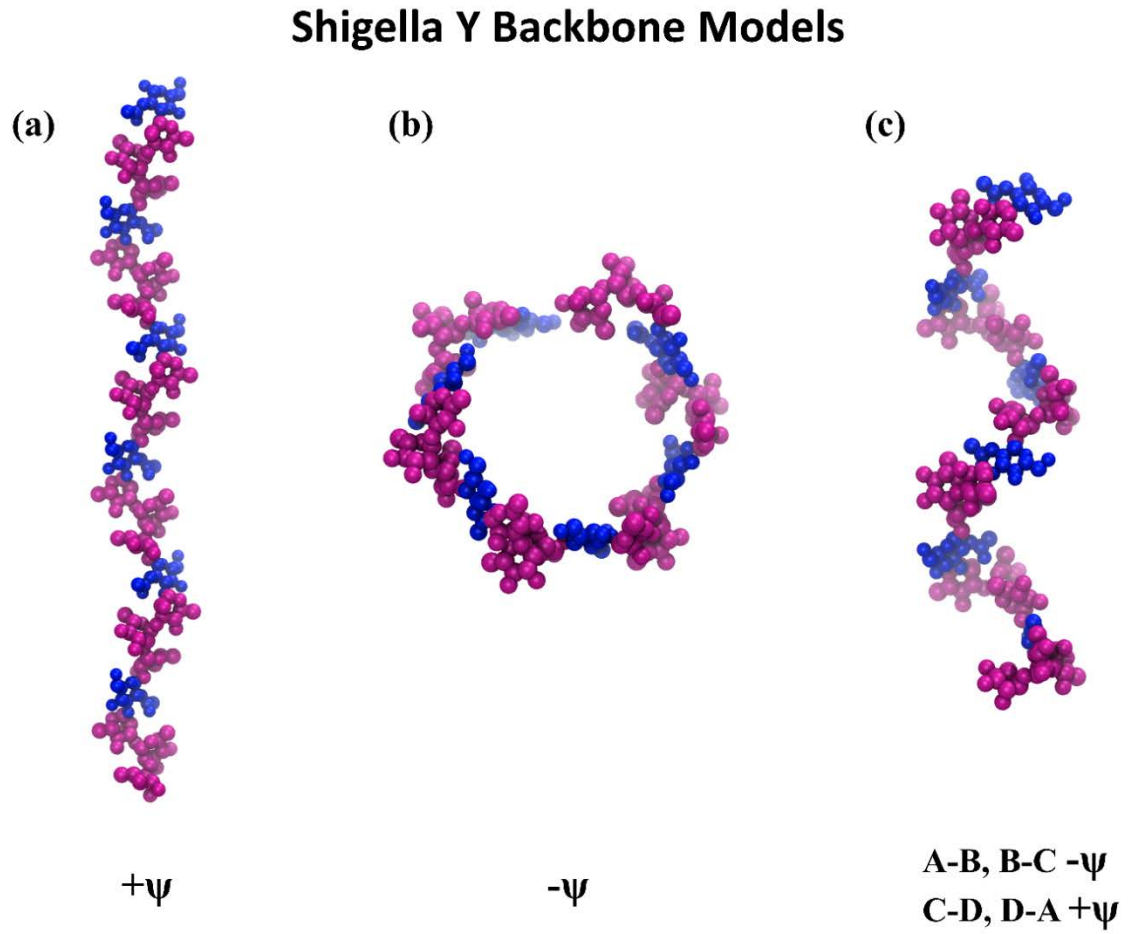
Abbreviations

MD: molecular dynamics; O-Ag: O-antigen; PMF: potential of mean force; RU: repeating unit

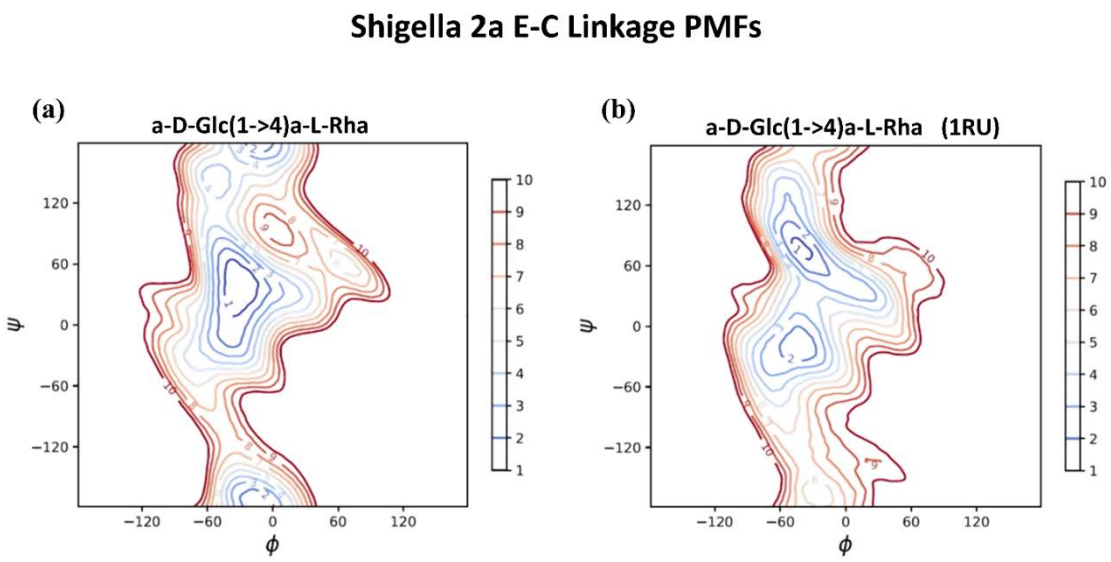
4.8 Supplementary information



Supplementary Figure S4.1: Block averaging analysis calculated for the time series of end to end distance (a) and radius of gyration (b) of the 6 RU O-Ags. The blocked standard error (BSE) reaches a plateau for all the O-Ags, indicating convergence of the simulation. Further analysis reveals correlation times consistently less than 45 ns – much less than the simulation time of 1000 ns, indicating sufficient sampling.



Supplementary Figure S4.2: The 6 RU static models of the serotype Y backbone built by extrapolation of the dominant linkage conformations in the simulations which demonstrate the extremes in the polysaccharide motion. (a) A model with all linkages in the +ψ orientation. This creates an extended conformation and corresponds to the extended conformations (cluster Y1) only observed for Y. (b) The 6 RU model with linkages in the -ψ orientation produces a rounded conformation which corresponds to the conformations of 2a-3Ac. (c) The helical model produced by -ψ orientations in the A-B and B-C linkages with +ψ orientations in the C-D and D-A linkages which corresponds to the conformations observed in 2b.



*Supplementary Figure S4.3: Vacuum PMF plots of the α-D-glucose-(1->4)-α-L-rhamnose linkage corresponding to the E-C linkages of 2a for: (a) the Glc*p*-(1→ 4)-α-L-Rha*p *disaccharide and (b) a full pentasaccharide 2a repeating unit. The presence of the adjacent rhamnose (residue B) sterically constrains the E-C linkage to two distinct conformations separated by an energy barrier of between 3-4 kcal.mol^{-1}.*

Chapter 5

Cross-protection from a *S. flexneri* multivalent vaccine including serotypes 2a and 3a

5.1 Foreword

The following article "*Molecular Modeling of the* Shigella flexneri *Serogroup 3 and 5 O-Antigens and Conformational Relationships for a Vaccine Containing Serotypes 2a and 3a*" expands the investigation of *S. flexneri* O-Ags to determine the prospect of cross-protection occurring from a proposed multivalent vaccine that includes serotypes 2a and 3a.

As the chemical structure of the *S. flexneri* O-Ags are fairly conserved between strains, a multivalent vaccine with just three strains of *S. flexneri* (serotypes 2a, 3a and 6) is proposed to provide protection against the majority of prevalent infectious strains. Together, serotypes 2a and 3a express shared antigenic epitopes with almost all remaining serotypes, except for serotype 6. As vaccine development is still ongoing, only limited cross-reactivity data is available. Therefore, molecular dynamics simulations can offer valuable insights to guide the investigation toward an optimal set of vaccine components for a multivalent *Shigella* vaccine.

In addition to the serogroup 2 O-Ags studied in the previous chapter, this study considers serogroup 3 as well as the non-vaccine serogroup 5. We identify three general heuristics for the conformational effects of different substitution patterns on the →2)-α-L-Rha*p*^{III}-(1→2)-α-L-Rha*p*^{II}-(1→3)-α-L-Rha*p*^{I}-(1→3)-β-D-Glc*p*NAc-(1→ O-Ag backbone (residues 'A-B-C-D'): first, substitution on rhamnose C reduces overall chain extension; second, substitution on rhamnose A induces predominantly helical conformations; third, substitution on rhamnose B has only a minor conformational effect. Further, our modeling indicates that broad cross-protection against non-vaccine strains is likely from vaccination with serotypes 2a and 3a, through shared C-curve and helical conformations. These results support and guide the development of an envisioned broad-coverage multivalent *Shigella* vaccine.

Molecular Modeling of the Shigella flexneri *Serogroup 3 and 5 O-Antigens and Conformational Relationships for a Vaccine Containing Serotypes 2a and 3a*

Jason Hlozek[a], Sara Owen[b], Neil Ravenscroft[a] and Michelle M. Kuttel[b]*

[a] Department of Chemistry, University of Cape Town, Rondebosch 7701, South Africa

[b] Department of Computer Science, University of Cape Town, Rondebosch 7701, South Africa

* Correspondence: Tel: +27 21 650 5107; E-mail address: mkuttel@cs.uct.ac.za (Michelle M. Kuttel)

Keywords: O-antigen; conformation; *Shigella flexneri*; molecular modeling; cross protection

5.2 Abstract

The pathogenic bacterium *Shigella flexneri* is a leading global cause of diarrheal disease. The O-antigen is the primary vaccine target and distinguishes the 30 serotypes reported. Except for serotype 6, all S. flexneri serotypes have a common backbone repeating unit (serotype Y), with variations in substitution creating the various serotypes. A quadrivalent vaccine containing serotypes 2a and 3a (as well as 6 and *S. sonnei*) is proposed to provide broad protection against non-vaccine *S. flexneri* serotypes through shared epitopes and conformations.

Here we model the O-antigen conformations of serogroups 3 and 5: a continuation of our ongoing systematic study *S. flexneri* O-antigens that began with serogroup 2. Our simulations show that *S. flexneri* serogroups 2, 3 and 5 all have flexible O-Ags, with substitutions of the backbone altering the chain conformations in different ways. Our analysis suggests three general heuristics for the effects of substitution on the Shigella O-Ag conformations: (1) substitution on rhamnose C reduces the extension of the O-Ag chain; (2) substitution at O-3 of rhamnose A restricts the O-Ag's to predominantly helical conformations, (3) substitution at O-3 of rhamnose B has only a slight effect on conformation. The common O-Ag conformations across serotypes identified in this work support the assumption that a quadrivalent vaccine containing serotypes 2a and 3a could provide coverage against *S. flexneri* serotype 3b and serogroup 5.

5.3 Introduction

Diarrheal diseases cause over 1.6 million deaths each year,[96] disproportionately affecting low-income regions[130] and young children.[95] *Shigella flexneri* is a leading cause of enteric infections, with no licensed vaccine currently available.[131] The increasing prevalence of antibiotic resistant strains[132-135] necessitates a broad coverage *Shigella* vaccine to prevent infection[131, 136] and reduce the global disease burden.[137, 138]

The structure of the *S. flexneri* O-antigen (O-Ag) repeating unit (RU) - the glycan component of the cell-surface lipopolysaccharide - classifies strains into approximately 30 serotypes and 7 serogroups.[100, 139] The O-Ag is the primary target of the host immune response[140, 141] and is a focus of current vaccine development.[101, 113] Except for serotype 6, all *S. flexneri* serotypes have the same backbone RU - serotype Y: →2)-α-L-Rha*p*^{III}-(1→2)-α-L-Rha*p*^{II}-(1→3)-α-L-Rha*p*^{I}-(1→3)-β-D-Glc*p*NAc-(1→ (Figure 5.1a). Serotypes are defined by the combination of type O-factor epitopes in the O-Ag RU (which distinguish the serogroups from each other) as well as the group O-factors determined by glucosylation, O-acetylation and phosphorylation substitution of the serotype Y backbone, which appear across serogroups.[139, 142] The similarities between *S. flexneri* O-Ags (shared backbone and group O-factors) suggest that some serotypes may cross-protect, enabling development of a broad-coverage vaccine with minimal valency. The proposed vaccine for *Shigella* based on the Global Enteric Multicenter Study (GEMS) consists of *S. flexneri* serotypes 2a, 3a, 6 and the single *S. sonnei* serotype, which is estimated to provide direct protection against 64% of *Shigella* strains causing infection in children in low-income areas, with cross-reactivity potentially extending this up to 88%.[98, 105, 107] Serotypes 2a and 3a were chosen as

vaccine components because both are prevalent causes of infection (ranked 1st and 4th, respectively) and together they express the group O-factor epitopes (group O-factor 6, 7,8 and 9) found on most remaining non-vaccine serotypes, allowing for potential broad cross-protection.

As cross-protection between antigens is expected to require both chemical and conformational similarity,[22] molecular modeling can provide insight into the potential for cross-protection between *S. flexneri* O-Ags. Early computational models indicated an extended conformation for serotype Y,[143] and a helix for serotype 5a.[127] However, MD simulations predicted a that serotype Y is highly flexible [144, 145] and that 3 RU of 12 *S. flexneri* O-Ags show similar conformations to each other.[146],[147]. Although *S. flexneri* expresses a heterogenous distribution of O-Ag chain lengths, 3 RU is considered sufficient to represent the O-antigen conformation.[113, 146]

We previously embarked on a systematic conformational study of all the Shigella O-Ags, beginning with the backbone (serotype Y) and the serogroup 2 O-Ags.[145] For the serotype 2a O-Ag (Figure 5.1b), we found that glucosylation on O-4 of rhamnose C (type O-factor II)[139] restricted the O-Ag to more compact conformations as compared to the highly flexible, unsubstituted serotype Y. Additional substitution on O-3 of rhamnose A produced more extended helical conformations, regardless of whether it was O-acetylation (serotype 2a, group O-factor 9); or glucosylation (serotype 2b; group O-factor 7,8; Figure 5.1c). This work indicated that an O-3-acetylated 2a O-Ag (expressing group O-factor 9) may provide stronger cross-protection against 2b (ranked 3rd in prevalence, expressing group O-factor 7,8) than 2a (which lacks both O-

factors) and may provide enhanced coverage of other serotypes expressing O-factor 9 (serotypes Y_1, $1a_1$, 1b, $5a_1$, 6, and $7a_1$).[139]

Here we report the next step of our study, modeling the conformations of serotypes 3a, 3b and serogroup 5. Serotype 3a is a component of the proposed quadrivalent vaccine[98, 105] as it is prevalent globally[98]. Serotype 3b is not included in the vaccine, but is prevalent in parts of Asia.[148, 149] In contrast, serogroup 5 has a relatively low incidence of disease; serotype 5a has been widely studied and is used as a reference strain, whereas 5b accounted for some cases in the GEMS report.[98]

All serotypes in serogroup 3 contain O-acetylation on O-2 of rhamnose C (group O-factor 6),[139] which is important for antibody recognition.[150, 151] Serotype 3a (Figure 5.1d) is defined by additional glucosylation on O-3 of rhamnose A (O-factor 7,8), with subtype $3a_1$ having partial O-acetylation on O-6 of GlcNAc D (O-factor 10, ~40%); neither of which are present in serotype 3b (Figure 5.1e).[100, 139] Serotype 5a (Figure 5.1f) is defined by glucosylation at position O-3 of rhamnose B (type O-factor V) with subtype $5a_1$ having partial O-3-acetylation on rhamnose A (group O-factor 9, ~35%).[139] The most common 5a strain for laboratory study (M90T) has no O-acetylation.[152, 153] Serotype 5b (Figure 5.1g) is glucosylated on O-3 of rhamnose A (O-factor 7,8) with no O-acetylation.

Some studies have indicated cross-reactivity between serogroups 2, 3 and 5. Serotypes 2a, 3b and 5a each react with group O-factor 3,4 antisera (associated backbone trisaccharide residues C-D-A)[100] and partial cross-reactivity with these strains is demonstrated from serotype 2a in human testing.[104] Serotypes 2b, 3a and 5b share glucosylation on O-3 of rhamnose A (group O-factor 7,8) and strong cross-reactivity is

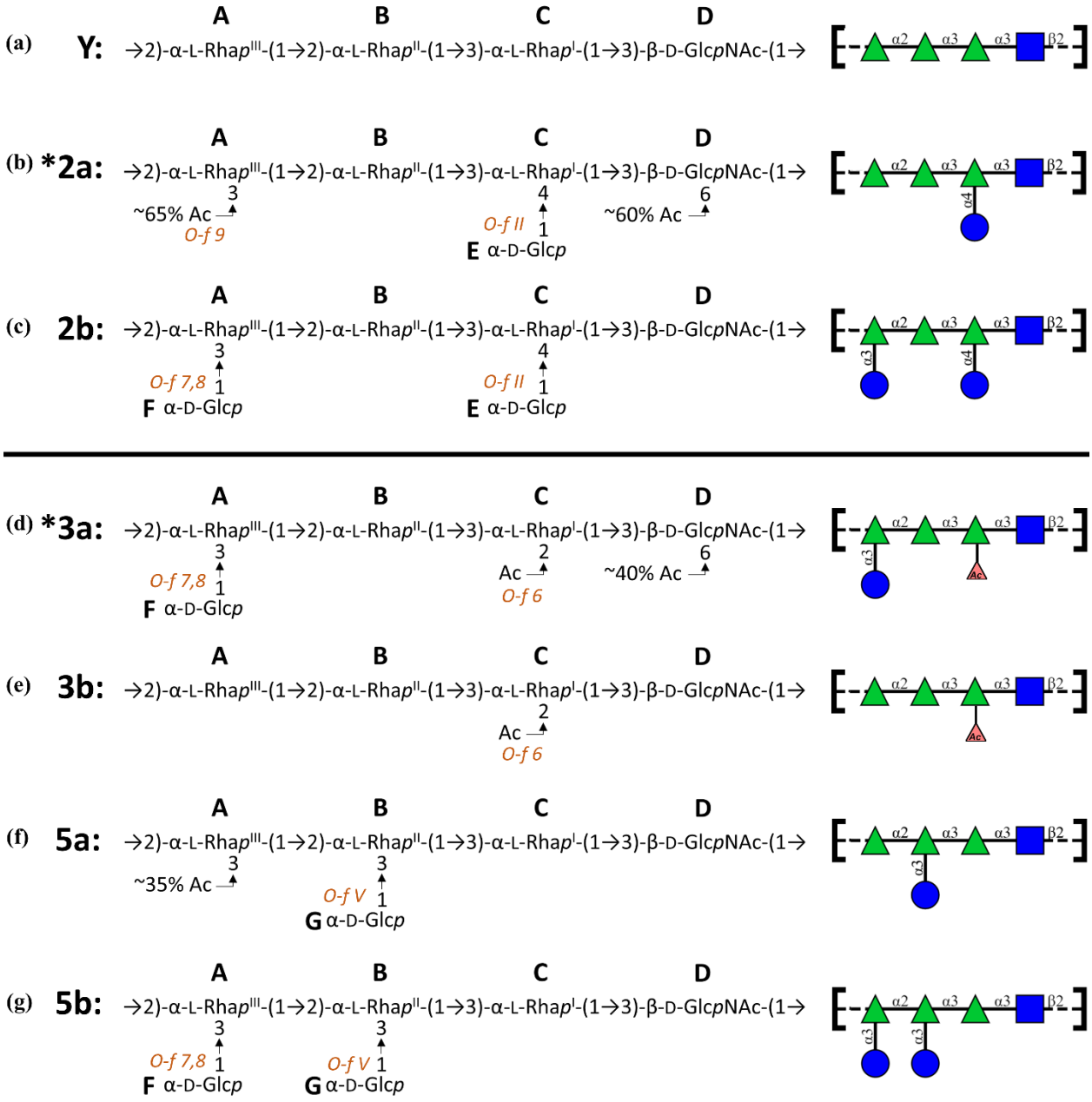


Figure 5.1: Line structures and diagrams of the Shigella flexneri O-antigen repeating units of serotypes (a) Y, (b) 2a, (c) 2b, (d) 3a, (e) 3b, (f) 5a, and (g) 5b. All serogroups share the serotype Y backbone and are distinguished by substitutions, which are labelled with the associated O-factors. Vaccine serotypes 2a and 3a are indicated with an asterisk. Schematic diagrams are depicted using the ESN symbol set[108] where: green triangle - Rha, blue square - GlcNAc, blue circle – Glc, red triangle – O-acetylation.

reported from a candidate 2a/3a vaccine against 2b and 5b in guinea pigs.[106] Therefore, a vaccine containing serotypes 2a and 3a (expressing group O-factors 6; 7,8 and 9) is suggested to elicit broad cross-protection against the remaining serogroup 2, 3 and 5 serotypes.[98]

Here we compare simulations of serogroups 3 and 5 with our recent work on the serogroup 2 O-Ags, contrasting the O-Ag behavior for O-acetylated serotype 2a (group O-factor 9) and 3a (group O-factors 6 and 7,8) with the non-vaccine serotype 3b (group O-factor 6) as well as serotypes 5a (type O-factor V) and 5b (group O-factor 7,8). With this large data set for comparison, we aim to broadly identify guiding heuristics for the conformational effect of substitutions on particular positions of the *Shigella* backbone, specifically substitutions on rhamnose C (O-factor II in serogroup 2 and O-factor 6 in serogroup 3); O-3 of rhamnose A (O-factor 9 in serotype 2a-3Ac; O-factor 7,8 in serotypes 2b, 3a and 5b); and O-3 of rhamnose B (O-factor V in serogroup 5). We assume that O-factors with a significant conformational effect should be represented in the vaccine serotypes to allow for broad coverage. Ultimately, we aim to determine whether the conformational findings from our computational modeling supports the assumption that a quadrivalent vaccine containing *S. flexneri* serotypes 2a and 3a (as well as 6 and *S. sonnei*) could provide broad coverage against *S. flexneri* serotype 3b and serogroup 5.

5.4 Materials and methods

The *S. flexneri* O-Ags have glycosidic linkages described by two torsion angles, ϕ and ψ, defined as: ϕ = H1-C1-O1-C_x' and ψ = C1-O1-C_x'-H_x'. These definitions are analogous to ϕ_H and ψ_H in IUPAC nomenclature and are consistent with our previous glycan

modeling.[1, 145] This work follows our established methodology for the computational study of glycan antigens. Potential of mean force (PMF) calculations for the disaccharide fragments of the O-Ag repeating unit indicate the global minima for each linkage, which are then used to build short 3 RU chains for initial 300 ns MD simulations in solution. The most populated linkage conformations from the 3 RU simulations are then used to construct starting structures for the simulations of the 6 RU chains.[2, 124, 154]

5.4.1 φ, ψ PMF calculations

The low-energy conformations of the glycosidic linkages were determined by calculating the potential of mean force (PMF) for rotation about the ϕ and ψ torsion angles of each disaccharide linkage. PMFs were calculated with the metadynamics algorithm[36] as implemented in NAMD.[155] The disaccharide PMFs were calculated in the gas-phase with the ϕ, ψ torsion angles set as collective variables.

5.4.2 Molecular dynamics

Simulations were run with the NAMD software package,[155] employing CUDA extensions to leverage graphics processing units for the calculation of long-range electrostatic potentials and non-bonded forces.[156] Glycans were modeled with the CHARMM36 additive force field for carbohydrates[31, 32] and explicit water molecules were represented with the TIP3P water model.[157]

Our in-house CarbBuilder software was used to build the glycan structures prior to simulation.[158] Initial 3 RU chains of the serogroup 3 and 5 O-Ags (not discussed here) were built with glycosidic linkage conformations set to the energy minimum of the

respective disaccharide PMFs. The RUs of the O-Ag chains modeled in this study are as follows with the serotype-defining moieties in bold:

3a: →2)-***[αDGlc(1→3)]***αLRha(1→2)-αLRha(1→3)-αLRha**2Ac**(1→3)-βDGlcNAc-(1→

3b: →2)-αLRha(1→2)-αLRha(1→3)-αLRha**2Ac**(1→3)-βDGlcNAc-(1→

5a: →2)-αLRha(1→2)-***[αDGlc(1→3)]***αLRha(1→3)-αLRha(1→3)-βDGlcNAc-(1→

5b: →2)-***[αDGlc(1→3)]***αLRha(1→2)-***[αDGlc(1→3)]***αLRha(1→3)-αLRha(1→3)-βDGlcNAc-(1→

The 3 RU simulations were run for 1,000 ns and 300 ns (for serogroup 3 and 5, respectively), with the most frequent torsion angles from these simulations used to build the initial conformations for the 6 RU chains. These conformations were then subjected to 10,000 steps of standard NAMD minimization in vacuum and subsequently placed into a cubic water box with the *solvate* command from the Visual Molecular Dynamics (VMD) package.[159] The cubic water boxes for the 6 RU structures had side lengths of 100 Å and 90 Å, respectively, for serogroup 3 and 5, and periodic boundary conditions were employed. The solvated structures were gradually heated through a protocol of 5 K incremental temperature reassignments between 10 K and 310 K, with 1,000 steps of NAMD minimization and 1,000 steps of MD after each temperature reassignment.

Equations of motion were integrated using the velocity-Verlet method[160] with a 1 fs step size. Molecular dynamics simulations were performed under the isothermal-isobaric (nPT) ensemble at a temperature of 310 K and maintained with a Langevin piston barostat[155] and Nose-Hoover thermostat – a hybridized method of the Nose-Hoover constant pressure method[161] with piston fluctuations controlled by Langevin

dynamics,[75] as implemented in NAMD. Particle mesh Ewald (PME) summation[76] was used for calculation of long-range electrostatics, with k = 0.20 $Å^{-1}$ and PME grid dimensions that were set equal to the periodic cell size. Non-bonded interactions were truncated at 15.0 Å and a switching function implemented between 12.0 Å and 15.0 Å. 1-4 interactions were not scaled, in accordance with CHARMM force field recommendations.

5.4.3 Block averaging analysis

Block averaging analysis is used to assess simulation convergence and is implemented with in-house Python scripts.[126] The block averaging analysis algorithm splits a simulation trajectory with *N* frames into a set of *M* "blocks" with a length of *n* frames, such that $N = M \times n$. Next, an average of a selected measurable (e.g. end-to-end distance) is calculated within each block. The block length (*n*) is slowly increased and, at each value of *n,* the set of block averages are recalculated. The standard deviation in the set of block averages, σ_n, is used to determine the blocked standard error (BSE) for each value of *n*. The simulation is indicated to be converged once the running estimate of the BSE asymptotes to a plateau, where the plateau represents the true standard error in the estimate of the mean. [35]

5.4.4 Data analysis

Simulations underwent 200 ns of equilibration followed by production runs of 1 µs and 2 µs (for serogroup 3 and 5, respectively). Snapshots of molecular conformations were taken at 25 ps intervals from the simulation trajectories. Inter-atomic distances and torsion angles were measured from VMD's Tcl console and graphical user interface, and statistical calculations were performed with in-house Python scripts. For all

saccharides, we defined the end-to-end distance, *r*, as the length from C-2 of rhamnose B at the non-reducing end of the chain and C-1 of rhamnose C at the reducing end, thus excluding the very flexible terminal sugar units.

Molecular conformations were visualized in VMD,[159] with carbohydrate rings highlighted by the PaperChain visualization algorithm.[162] Before conformational clustering, the trajectory snapshots were aligned on the ring atoms of the central 'C-D-A-B' fragment between RU3 and RU4 – a frame-shifted full repeating unit to account for the variability in each linkage across the different serotypes. The most common chain conformations are determined by clustering the simulation snapshots into families with relative occupancies. We cluster the central 4 RU of each 6 RU chain, as the terminal repeating units are less representative of the native O-Ag backbones. VMD's internal *cluster* command was employed to calculate the conformational clusters in the production runs with an RMSD fit of the non-hydrogen atoms in the central 4 RUs with a cut-off of 5.5 Å. Clusters comprising less than 5% of the simulation were excluded. The conformations of the previously published serotype Y and serogroup 2 simulations[145] were recalculated under these criteria for a fair comparison between the serogroups.

5.5 Results

We begin our analysis of the simulation data with a broad comparison of the O-Ag chain extension and flexibility of *S. flexneri* serogroup 2 with serogroups 3 and 5; then we analyze the dominant backbone conformation of each O-Ag; and finally we explore the conformational effects of the glucosylation and O-acetylation on the orientations of the backbone glycosidic linkages.

5.5.1 Simulation convergence

We used block averaging analysis[35, 126] of two metrics of chain flexibility — the end-to-end distance, r, and the radius of gyration, R_g, — to assess the convergence of the MD simulations. Convergence is indicated by the plots of the blocked standard error (BSE) for both r and R_g (shown in Figure S5.1a and S5.1b) reaching a plateau. The asymptote of the BSE plot represents the true standard error in the measured variable, which can be used to approximate a correlation time for the simulation. The range of correlation times from 18 ns to 104 ns indicate that the 200 ns equilibration period is sufficient for all the O-Ags in this study. Further analysis reveals that the number of statistically independent samples in each simulation is much greater than 1 (49, 58, 88 and 21 for 3a, 3b, 5a and 5b, respectively) - as recommended for a converged trajectory.[35] Therefore, block averaging analysis indicates that the longer production runs (1,000 ns for serogroup 3 and 2,000 ns for serogroup 5) provide sufficient sampling of the conformational space. The simulations of the more flexible serogroup 5 O-Ags were extended to 2 μs to ensure that convergence was achieved.

5.5.2 O-Ag flexibility

The fluctuation in r over the course of a simulation is a simple measure of molecular extension and flexibility for the *S. flexneri* O-Ag's. Here we define *r* as the distance between C-2 of rhamnose B in RU1 and C-1 of rhamnose C in RU6 (Figure 5.2a). Figure 5.2 compares the *r* time series and corresponding histograms for the simulations of the serotype Y backbone and the serogroup 2 O-Ags previously published[145] with simulations of serogroup 3 (Figure 5.2f-5.2g) and serogroup 5 (Figure 5.2h-5.2i).

For a first broad comparison of the O-Ags, a scan of the graphs in Figure 5.2 quickly reveals that the unsubstituted Y backbone (Figure 5.2b) is by far the most flexible of the O-Ags, showing the greatest range of *r* values (σ ≈ 16 Å), whereas the serogroup 2 O-Ags (Figure 5.2c-5.2e) are the least flexible (σ ≈ 8-9 Å). Serogroups 3 and 5 fall between these two extremes, with serogroup 3 being somewhat less flexible (σ ≈ 13 Å for 3a, 14 Å for 3b) than serogroup 5 (σ ≈ 14 Å for 5a, 15 Å for 5b). Moreover, the distributions of *r* vary considerably across the O-Ags. The flexible serogroup Y has a bimodal distribution, in stark contrast to the well-defined tight distributions shown for serogroup 2.

The graphs of *r* reveal that the effect of glucosylation on the conformation and dynamics of the O-Ag chain depends on the glucosylation position: compare the graph of serotype 2a expressing O-factor II (O-4 glucosylation on rhamnose C, Fig 5.2c) with 5a expressing O-factor V (O-3 glucosylation on rhamnose B, Figure 5.2h). In particular, the O-factor II glucosylation that defines serogroup 2 has a very dramatic effect on *r*, reducing the overall extension and flexibility of the chain. In contrast, the O-factor V glucosylation that defines serogroup 5 has a less obvious effect on *r*, slightly increasing the average chain extension. Finally, O-factor 7,8 (O-3 glucosylation on rhamnose A; serotypes 2b, 3a and 5b) appears to narrow the distribution of *r*, making the O-Ag chains more conformationally defined. For example, there is a clear difference in the shape of the *r* distribution for serotype 3a (Figure 5.2f) as compared to 3b (Figure 5.2g): serotype 3a has a unimodal distribution of r (mean 45 Å, σ ≈ 13 Å) whereas in 3b (which lacks O-factor 7,8) *r* is shifted to smaller values and has a right-skewed distribution with a peak at 25 Å (mean 29 Å, σ ≈ 14 Å). This significant shift in *r*

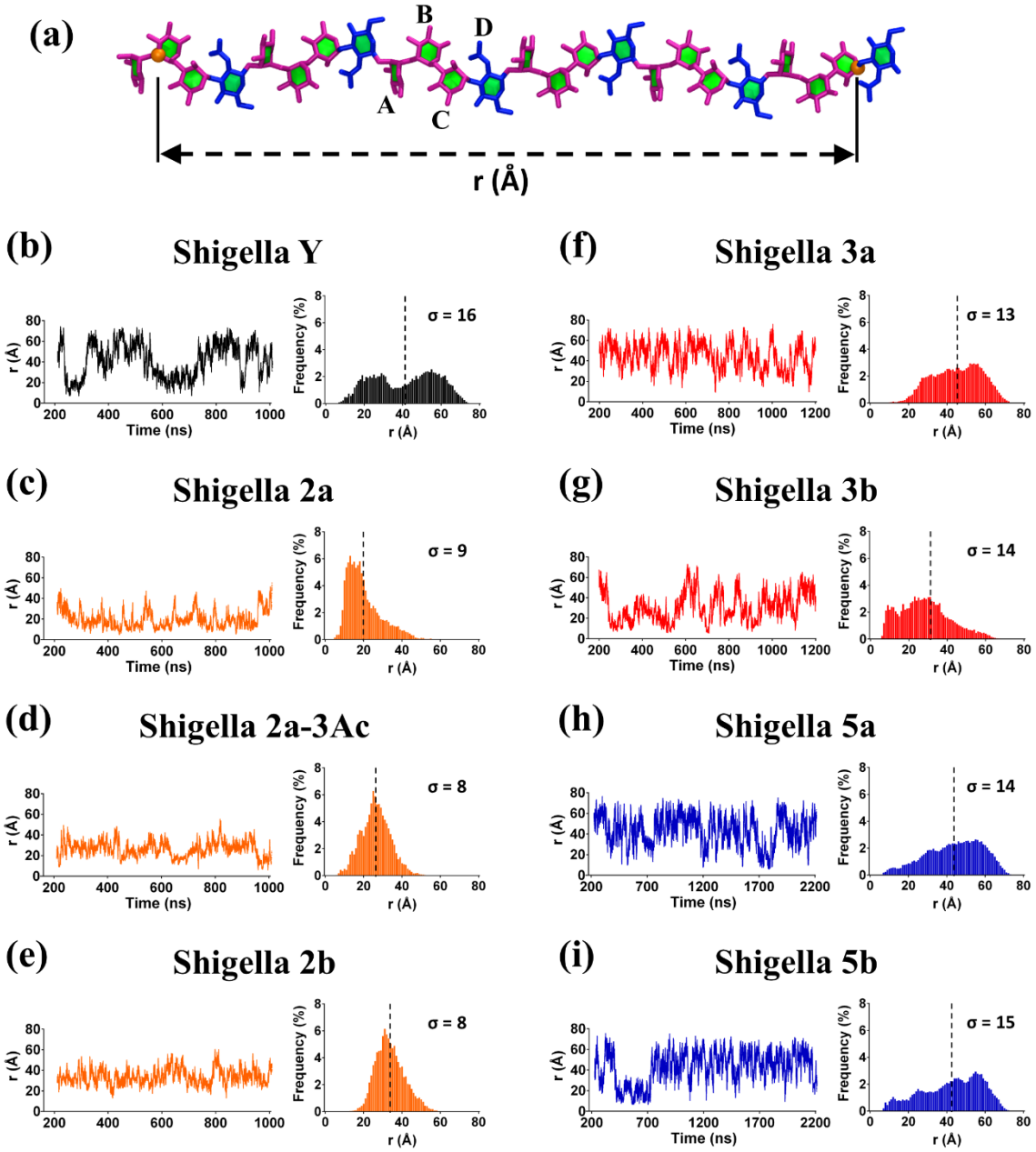


Figure 5.2: Comparison of the r time series and corresponding histograms for 6 RU simulations of the modeled S. flexneri *O-Ags. (a) A 6 RU model of the serotype Y O-Ag depicted with the end-to-end distance,* r*; rhamnose is colored pink and N-acetyl-glucosamine blue. The* r *time series (left column for each serotype) and corresponding distribution (right column for each serotype) are shown for (b) Y, (c) 2a, (d) O-3 acetylated 2a, (e) 2b, (f) 3a, (g) 3b, (h) 5a and (i) 5b. The mean for each histogram distribution is depicted with a dashed line and the corresponding standard deviation is indicated.*

distribution within serogroup 3 indicates a substantial conformational difference between the 3a and 3b O-Ags, and hence a significant conformational effect for O-factor 7,8. We have previously observed the same effect in serogroup 2 and it can be observed within serogroup 5, which is the least conformationally defined of the four serogroups. Serogroup 5 is the most similar to the Y backbone, but does not show the same clear bimodal distribution. The *r* distributions for serotype 5b (expressing O-factor 7,8) show a slight shift to more extended conformations (Figure 5.2i) as compared to 5a (Figure 5.2h). However, the similarity of the *r* distributions for 5a and 5b suggest similar conformational behavior for both serotypes.

Further, a comparison of the *r* histograms can reveal the general effects of O-acetylation on the O-Ag backbone. We have previously observed that O-factor 9 (O-acetylation at O-3 of rhamnose A, serotype 2a-3Ac) has a similar conformational effect to glucosylation at this position (O-factor 7,8, serotype 2b), reducing the range of *r*. The range restriction for O-acetylation is not as dramatic as with glucosylation, which is expected from the relative size of the substituents (Ac = 43 g/mol and Glc = 179 g/mol). More surprisingly, we now see that O-factor 6 (O-acetylation at O-2 of rhamnose C, serogroup 3) has a similar effect restrictive effect on *r,* as can be seen in a comparison of serotype 3b (Figure 5.2g) with the backbone serotype Y (Fig 5.2b).

A general comparison of the *r* histograms suggests three broad heuristics for the effects of substitution on the backbone (serotype Y) conformation of the *Shigella* O-Ags. First, substitution at rhamnose C (O-factor II in serogroup 2 and O-factor 6 in serogroup 3) has the most impact in reducing the flexibility and extension of the O-Ag chain, with glucosylation (2a) having a greater impact than O-acetylation (3b). The

impact of substitution on rhamnose C on O-Ag conformation is supported by the fact that O-factor II defines serogroup 2 and O-factor 6 is present in all of serogroup 3. Second, any substitution on O-3 of rhamnose A (O-factor 9 in serotype 2a-3Ac; O-factor 7,8 in serotypes 2b, 3a and 5b) shifts the O-Ag to more extended conformations. Third, O-factor V (substitution at O-3 of rhamnose B, serogroup 5) does not have a significant effect on chain conformation. Our analysis suggests that, at a first approximation, these three heuristics are additive. To test and refine these rules of thumb, as well as investigate the potential for cross-reactivity between serotypes, we now perform a detailed comparison of the chain conformations for all O-Ags.

5.5.3 O-Ag chain conformations

Figure 5.3 compares the conformational families for serotype Y (Figure 5.3a) and the serogroup 2 O-Ags (Figure 5.3b-d) with the serogroup 3 (Figure 5.3e and f) and serogroup 5 (Figure 5.3g and h) O-Ags. As discussed in prior work,[145] the flexible Y backbone transitions between extended conformations (Figure 5.3a Y-1, Y-4 and Y-6) to more curved arrangements of the chain (Figure 5.3a Y-2, Y-3 and Y-5). Further, we showed that O-factor II (O-4 glucosylation at rhamnose C, serogroup 2) has a dramatic effect on the chain conformations, removing the extended conformations and restricting the O-Ag to a wide range of "C-curves" (Figure 5.3b). Additional substitution on O-3 of rhamnose A - whether O-acetylation in 2a (O-factor 9) or glucosylation in 2b (O-factor 7,8) was then shown to further restrict the chain and induce helical conformations (compare Figure 5.3c and d).

In common with serogroup 2, serogroup 3 is also substituted at rhamnose C, albeit with a smaller O-acetyl group at position 2 (O-factor 6). A comparison of the dominant

conformations of 2a (Figure 5.3b) and 3b (Figure 5.3f) reveals that O-factors 6 and II have a similar effect on the backbone. For the 3b O-Ag, O-2-acetylation on rhamnose C removes most of the extended conformations of the chain and restricts the O-Ag to a range of C-curve conformations (3b-1) and other folded conformations of the chain (3b-2, 3b-4 to 3b-7). The dominant curved chain conformation for 2a (2a-1, 31%) and 3b (3b-1, 11%) are remarkably similar, indicating a similar conformational effect of substitution at rhamnose C. However, the smaller O-acetyl substituent means that the 3b chain remains more flexible than the 2a and has a minor helical conformational family (3b-3, 7%).

As for serotype 2b, serotype 3a is substituted at rhamnose C as well as rhamnose A (O-factor 7,8). This combination of substitutions has a similar effect on the conformation of the 3a O-Ag (Figure 5.3e) as it does on 2b (Figure 5.3d), producing a dominant helical conformation with 3 RU per turn and an average pitch of 29 Å (3a-1, 11%). However, 3a remains more flexible than 2b and can adopt a wide range of helical conformations (3a-3, 3a-5), as well as partially extended chains (3a-2), fully extended chains (3a-4) and S-bends (3a-6). Therefore, the serogroup 3 O-Ags follow a similar trend to serogroup 2[145]: substitution on rhamnose C restrict the chains to curved conformations for both serotype 2a and 3b (Figure 5.3b and 5.3f) and additional substitution on O-3 of rhamnose A shifts the conformations towards helices for serotypes 2a-3Ac, 2b and 3a (Figure 5.3c, 5.3d and 5.3e). Further, the axially orientated O-acetyl groups of serogroup 3 are readily accessible for antibody binding in both serogroup 3 O-Ags, which supports the reported immunodominance of O-factor 6.[150]

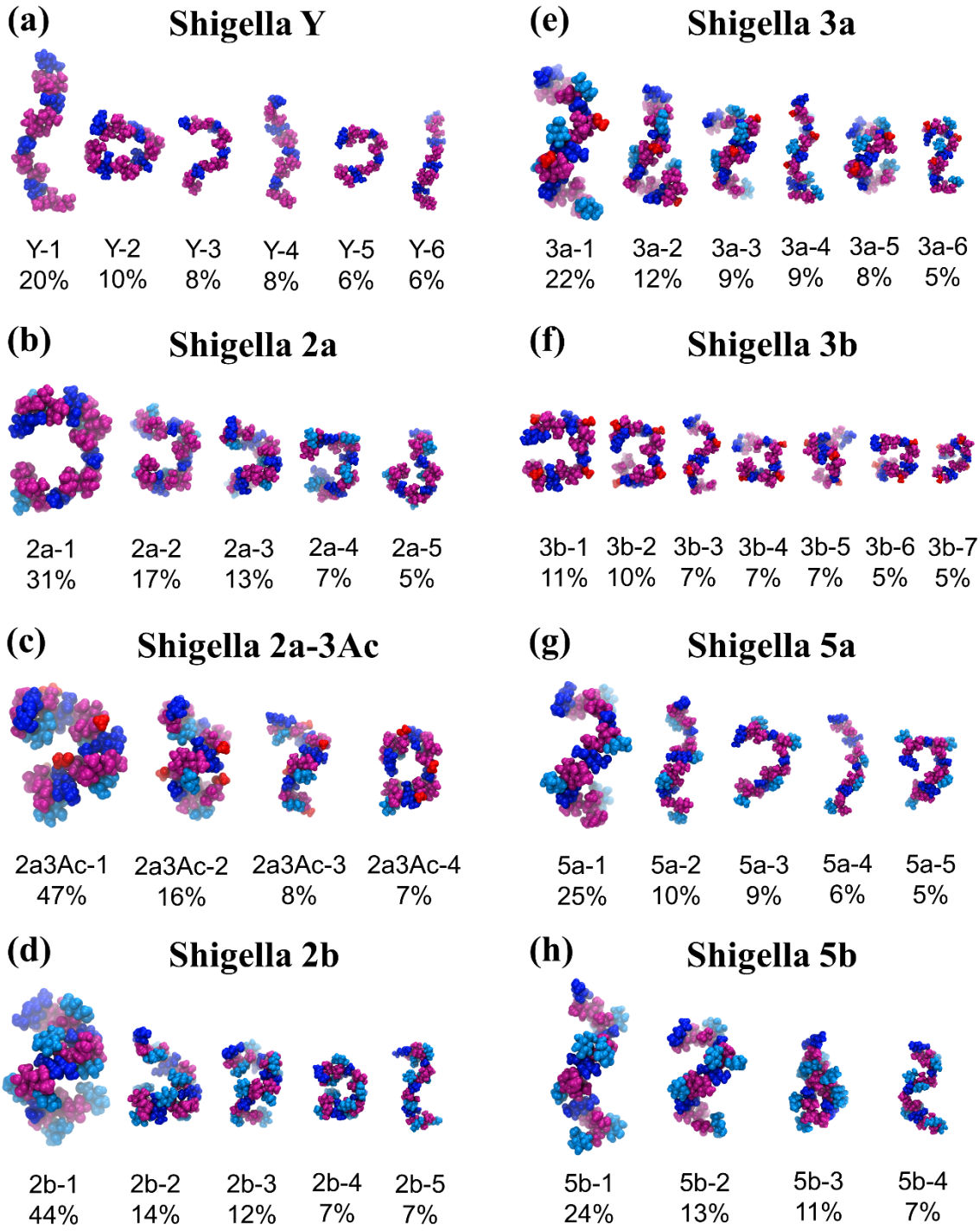


Figure 5.3: Conformational families of the central 4 RU of the 6 RU chains for (a) Y, (b) 2a, (c) O-3-acetylated 2a, (d) 2b, (e) 3a, (f) 3b, (g) 5a, and (h) 5b. Relative occupancies in the simulations (excluding the initial 200 ns) are indicated as percentages. Clusters of less than 5% are not shown. The sugars are colored: pink for Rha, dark-blue for GlcNAc, cyan for Glc side chains and red for O-acetyl groups.

A comparison of the primary conformations of serogroup 5 with the backbone shows the effect of O-factor V (glucosylation at O-3 of rhamnose B) on the chain conformation. Relative to the backbone (Figure 5.3a), the 5a O-Ag shows an increase in the prevalence of elongated helices (Figure 5.3g): the dominant conformation (5a-1, 25%) is a right-handed helix with 3 RU per turn, in agreement with an early helical model prediction.[127] This extended helical structure (pitch of 30 Å) has significant flexibility - the helix encompasses just 25% of the simulation and frequently unwinds to extended chains (5a-2 and 5a-4) as well as C-curve conformations (5a-3 and 5a-5) that are also present in the backbone (e.g. Y-4 and Y-5). Therefore, O-factor V has only a slight impact on the backbone conformation. However, the glucose side chains (colored cyan in Figure 5.3g and h) are exposed for antibody binding.

Serogroup 5b adds O-factor 7,8 (glucosylation on O-3 of rhamnose A) which, according to our heuristic hypothesis, should have a similar effect as in serogroup 2, increasing the dominance of helices in the 5b O-Ag conformations. This is in fact the case: although the 5b O-Ag has a similar dominant conformational family (Figure 5.3h, 5b-1, 24%) to 5a, the more minor C-curve conformations in 5a are replaced with helices in 5b (5b-2, 13%; 5b-4, 7%). The chain also has a unique tight hairpin bend conformation (5a-3, 11%) corresponding to the short *r* values adopted early in the 5b *r* time series (Figure 5.2i).

In summary, conformational analysis suggest refinement of our proposed three broad heuristics for the effects of substitution on the backbone conformation of the *Shigella* O-Ags, as follows. First, substitution at rhamnose C (present in serogroups 2 and 3) restricts the O-Ag chain to predominantly curved conformations, with a larger

substituent (e.g. O-4 glucosylation in serogroup 2) having a greater effect than a smaller one (e.g. O-2 O-acetylation in serogroup 3). Second, additional substitution on O-3 of rhamnose A (be it O-acetylation or glucosylation: 2a-3Ac, 2b, 3a, 5b) restricts the O-Ag to helical conformations. Third, substitution at O-3 of rhamnose B (O-factor V, serogroup 5) has only a slight effect on conformation, shifting the backbone to somewhat more extended O-Ag conformations. To explain the origin of these general effects, we now investigate the impact of the substitutions on the constituent glycosidic linkages in the *S. flexneri* O-Ag repeating unit.

5.5.4 O-Ag glycosidic linkage conformations

As carbohydrate rings have fairly constrained chair conformations, chain flexibility in the *S. flexneri* O-Ags arises principally from rotations about glycosidic linkages, which are commonly measured via the ϕ and ψ torsion angles. Ring substitutions can increase or, more commonly, decrease, the range of motion for a glycosidic linkage. The range of motion for each of the glycosidic linkages in the serogroup 3 and 5 O-Ag RUs over the course of the simulations is shown in Figure 5.4 with scatter plot heatmaps of the ϕ, ψ torsion angle distribution over the course of the simulations. Fragments of the O-Ag backbone showing the relative arrangements of the sidechains, the N-acetyl (blue) and the O-acetyl (red) substituents for each of the serotypes are shown in Figure 5.5.

For the Y-backbone (Figure 5.4a), the ϕ torsion angle for all linkages is restricted to a narrow range of values around ϕ ≈ 40°, whereas the ψ torsion angle is more flexible with two primary conformations at ψ ≈ 10° and ψ ≈ -35° (hereafter referred to as +ψ and –ψ, respectively). The D-A linkage is the most constrained, having the narrowest

range for psi, because the close proximity of the N-acetyl group to this β-D-Glc*p*NAc-(1→2)-α-L-Rha*p*^{III} linkage restricts ψ rotations. The backbone torsion angle conformations are consistent with the scatter plots of the ϕ, ψ linkages from short (60 ns) simulations of 3 RU chains.[146] Further, the estimates of key NOE distances (by r^6 averaging) are in good agreement with NMR NOE measurements for the native LPSs of serotype 3a[146] (Supp. Table S5.2) and 5a[127] (Supp. Table S5.3), providing validation for our MD simulations.[163]

Comparison of the heatmaps for substituted O-Ags with the backbone (serotype Y) maps allows a closer identification of the specific effect of substitutions on the O-Ag local chain flexibility and dynamics.

For the first heuristic, we have found that substitution at rhamnose C (serogroups 2 and 3) restricts the extension of the O-Ag chain to predominantly curved conformations. Comparison of the ϕ, ψ heatmaps for the Y backbone (Figure 5.4a) with those for serogroup 2 (Figure 5.4b-d) shows that the B-C glycosidic linkage (α-L-Rha*p*^{II}-(1→3)-α-L-Rha*p*^{I}) is considerably restricted in serogroup 2 as compared to the unsubstituted backbone. Serogroup 2 is glucosylated at O-4 of rhamnose C (O-factor II); steric hindrance by this glucose side chain restricts the range of freedom in the neighboring ψ torsion angle of the α-(1→3) B-C linkage – see Figure 5.5a-c. Serogroup 3 (Figure 5.4e and f) shows a lesser restraint on the B-C linkage, but the α-(1→3) C-D linkage is also restricted relative to the backbone. Serotype 3 is O-2-acetylated on rhamnose C (O-factor 6), and the O-acetyl group is in close proximity to the N-acetyl group of D (β-D-Glc*p*NAc), as shown in Figure 5.5d and e. Transient hydrogen bonding of the carbonyl oxygen with the hydrogens on the neighboring N-acetyl amplify the

restrictive effect of the O-acetylation: the C-D linkage for serotype 3b is constrained with partial access to –ψ conformations (mean ≈ -12°) and glucosylation on rhamnose A (O-factor 7,8) further limits the C-D linkage to +ψ orientations for serotype 3a (mean ≈ -6°). It is interesting that a similar restriction in either of these α-L-Rha*p*-(1→3) linkages (B-C and C-D) to –ψ angles has a similar conformational effect for the serogroup 2 and 3 O-Ags, reducing the extension of the O-Ag chain. In contrast, the serogroup 5 O-Ag B-C linkages remain unconstrained, adopting predominantly +ψ orientations (mean ≈ 7°) that result in a greater prevalence of extended structures despite glucosylation on O-3 of rhamnose B (O-factor V). This accounts for the more extended helical conformation in our molecular dynamics simulations as compared to the static model which was built with –ψ B-C orientations.[127]

In contrast, the α-L-Rha*p*^{III}-(1→2)-α-L-Rha*p*^{II} linkage (A-B) is flexible across all serogroup 2, 3 and 5 O-Ags (Figure 5.4, first column), significantly contributing to the conformational flexibility of the O-Ags. For serogroup 2, the linkage shifts to favor –ψ orientations (increasing the prevalence of C-curve conformations) while the serogroup 3 A-B linkages remain largely unchanged compared to the backbone (serotype Y). Inspection of the backbone fragments in Figure 5.5 shows that the glucose substituent on O-4 of rhamnose C is in closer proximity to the A-B linkage than the less bulky O-acetyl substituent at the O-2 position. For serogroup 5, the adjacent glucose side-chain on O-3 of rhamnose B does impose a small steric hindrance to the A-B linkage, possibly contributing to extension of the O-Ag chains in this serogroup.

For our second heuristic, we found that further substitution on O-3 of rhamnose A shifts the O-Ag to extended helical conformations. Comparison of the heatmaps for

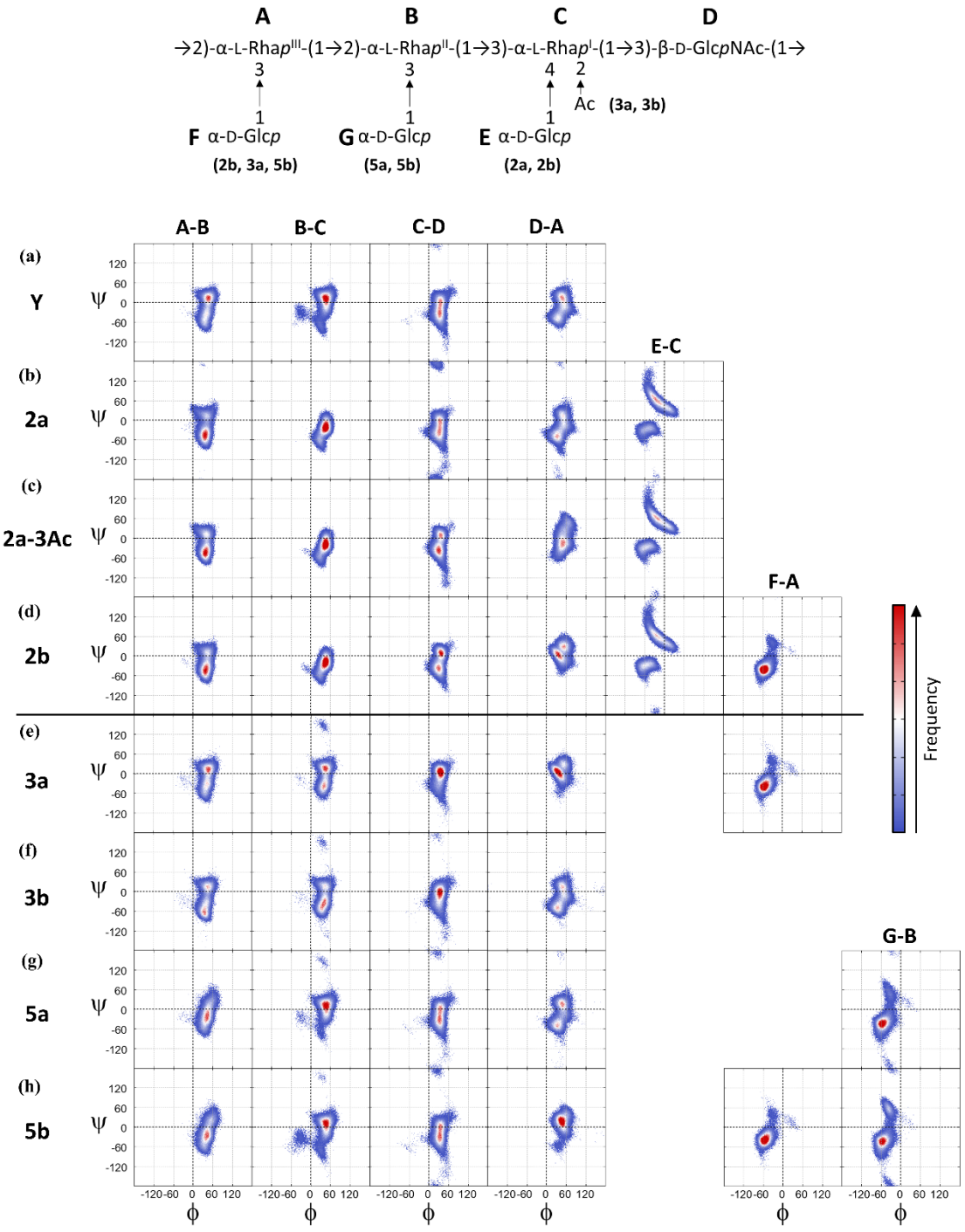


Figure 5.4: Heat map representations of scatter plots for the φ, ψ torsion angles of each glycosidic linkage for serotypes (a) Y, (b) 2a, (c) 2a-3Ac, (d) compared to serotypes (e) 3a, (f) 3b, (g) 5a and (h) 5b in this study. The heat maps combine the points from both central repeating units (RU 3 and RU 4) to broadly sample backbone behavior. The color scale to the right indicates the relative occupancy of the torsion angles during the simulations.

2b (Figure 5.4d), 3a (Figure 5.4e) and 5b (Figure 5.4h) with the serotype Y backbone shows that glucosylation at this position (O-factor 7,8) significantly reduces the range of motion for the ψ torsion angle of the D-A β-D-Glc*p*NAc-(1 →2)-α-L-Rha*p*^{I} linkage from two dominant conformations at ψ ≈ 15° (+ψ) and ψ ≈ -45° (-ψ) to a single dominant conformation at ψ ≈ 12°. In combination with -ψ orientations for the A-B and B-C linkages, this induces a primary helical conformation for serotype 3a. The glucose side-chain (residue F) is in close proximity to the N-acetyl group of the adjacent D residue **(**β-D-Glc*p*NAc), as shown in Figure 5.5c and d. Steric clashes between the glucose substituent and the N-acetyl group hinders rotation around the D-A linkage and reduces chain flexibility. The α-D-Glc*p*-(1→3)-α-L-Rha*p*^{III} side chain (F-A, Figure 5.4, fifth column) is in turn also restricted by the N-acetyl moiety with a single conformation at ϕ, ψ ≈ -52°, -36°, which is in agreement with NMR measurements for short serotype 3a fragments that predict a –ψ orientation.[150]

O-acetylation at this position (O-factor 9, serotype 2a-3Ac, Figure 5.4c) has a lesser, but similar, restriction on the range of rotation. Steric clashes between these two groups explain the large conformational restriction produced by glucosylation (serotypes 2b and 3a) as well as O-acetylation (2a-3Ac, Figure 5.5b) on rhamnose A. The conformational effects of this restraint are to increase the incidence of more regular helical structures in these serotypes.

Finally, for the third heuristic, we found that substitution at O-3 of rhamnose B (O-factor V, serogroup 5) has only a slight effect on conformation, shifting the backbone to somewhat more extended O-Ag conformations. Comparison of the heatmaps for serogroup 5a (Figure 5.4g) with the backbone shows that glucosylation at this position

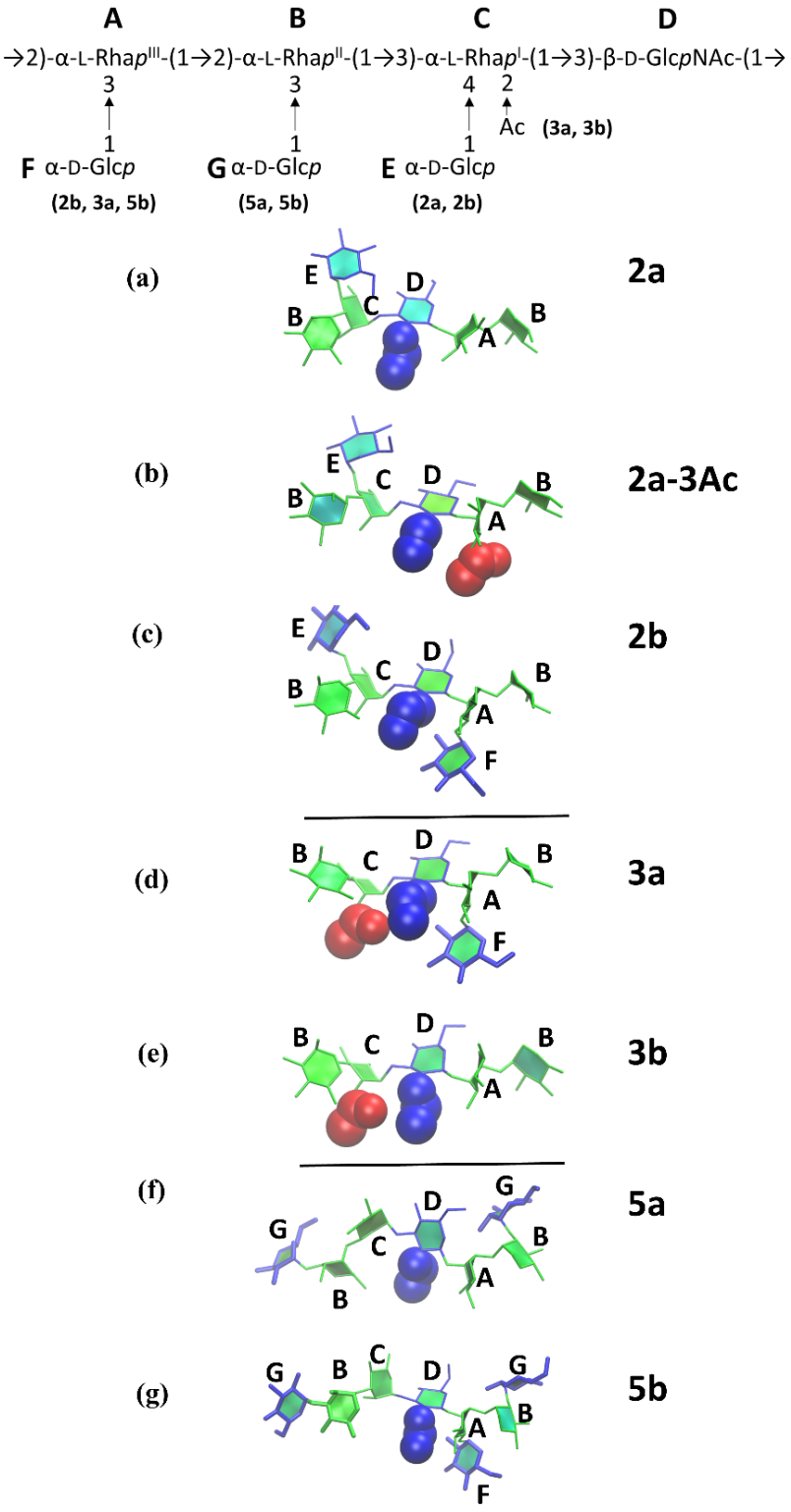


Figure 5.5: Fragments of the backbone showing the relative arrangements of the sidechains, the N-acetyl (blue) and the O-acetyl (red) substituents for serotypes (a) 2a, (b) 2a-3Ac, (d) 3a, (e) 3b, (f) 5a and (g) 5b. Glucose rings are colored blue, rhamnose green.

has little effect on the backbone A-B and B-C linkages: the glucose side chains do not interfere with rotations about the bonds (Figure 5.5f and g).

5.6 Discussion

Our simulations show that *S. flexneri* serogroups 2, 3 and 5 all have very flexible O-Ags. However, substitutions of the backbone residues limit the range and distribution of chain conformations in different ways. Our analysis has suggested three broad heuristics for the effects of substitution on the backbone conformation of the *Shigella* O-Ags: (1) substitution on rhamnose C has the greatest impact on restricting the extension and conformational range of the O-Ag's; (2) substitution at O-3 of rhamnose A (2a-3Ac, 2b, 3a, 5b) also has a strong impact, restricting the O-Ag's to predominantly helical conformations; (3) substitution at O-3 of rhamnose B (serogroup 5) has only a slight effect on conformation.

Can this conformational analysis give some insight into whether a quadrivalent vaccine containing *S. flexneri* serotypes 2a, 3a (as well as 6 and *S. sonnei)* could provide broad coverage against *S. flexneri* serotypes 3b, 5a and 5b? The factors that lead to cross-protection between O-Ags are not well understood. However, an assumption that similar O-Ag conformations is a necessary (if not sufficient) criterion for cross-protection between O-Ag's seems reasonable. However, immunodominant substitutions that change the binding surface (but perhaps not the conformation) may be confounding factors. On a conformational basis, we postulate that the two substitutions that produce the greatest conformational effects should be represented in the vaccine serotypes. Therefore, the vaccine should contain serotypes with substitutions on rhamnose C (O-factors II and 6) as well as rhamnose A (O-factors 9

and 7,8). Inclusion of an additional serotype with substitution at rhamnose B (O-factor V) seems less likely to be necessary. On the basis of this argument, the 2a-3Ac serotype containing both O-factors II and 9 would seem to be sufficient, whereas 2a (only substituted on rhamnose C) would not.

Cross-protection within serogroup 2 (2a-3Ac and 2b) seems likely due to the similar helical conformations of the serotypes within the group, and cross-protection with the helices in serogroup 3 and serogroup 5 may be possible. However, this conformational argument does not consider serogroup 3's immunodominant O-2-acetylation on rhamnose C (O-factor 6),[150, 151] which is a strong basis for including this serotype. Further reasons to include 3a are the prevalence of 3a infection, the lack of expected cross-protection from 2a[164-166] and potential cross-protection by 3a against 2b arising from the shared glucosylation of rhamnose A (O-factor 7,8). Although serotype 3b is conformationally more similar to the non-acetylated 2a chain, the minor C-curve and helical conformations of 3b may allow for partial cross-reactivity from a 2a-3Ac vaccine component. Further, serotypes 3a and 3b share minor conformational families, which may be sufficient to elicit cross-reactivity. Furthermore, serogroup 3a may provide cross-reactivity with other disease-causing serogroups: serotypes 1b and 4b express O-factor 6 and serotype X expresses O-factor 7,8.

5.7 Conclusion

On the basis of conformational similarity, we suggest that the inclusion of serogroup 5 is not necessary in the vaccine, as serotype 5a shares similar helical structures with the 2a-3Ac chain. Further, the partial O-acetylation on O-3 of rhamnose A (O-factor 9) for serotype 5a could provide cross-reactivity with an O-acetylated serotype 2a vaccine,

although the exposure of the O-factor V glucosylation for antibody binding may be a confounding factor. In future work, we will investigate the O-Ags of the next most prevalent serogroups identified by the GEMS report – serogroups 6 and 1 – to allow for further development of our heuristics for the conformations of the *S. flexneri* O-antigens.

Supplementary Materials: The following are available online at www.mdpi.com/, Figure S5.1: Blocked standard error calculations for determining the extent of simulation convergence, Table S5.2: Comparison of serotype 3a H^1 distances to NMR measurements and previous MD study, Table S5.3: Comparison of serotype 5a H^1 distances to NMR measurements and previous MD study.

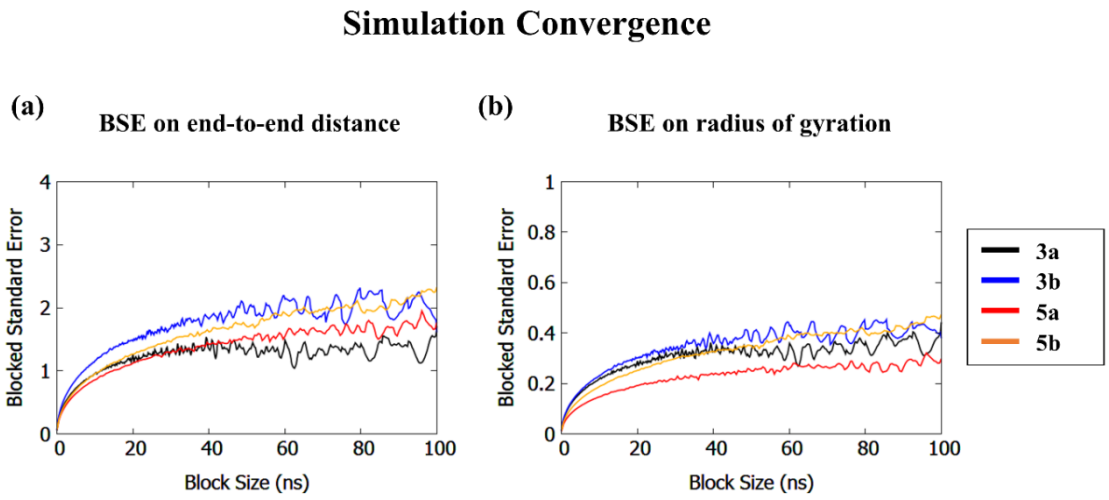
Author Contributions: Conceptualization, Neil Ravenscroft and Michelle M. Kuttel; Funding acquisition, Neil Ravenscroft and Michelle M. Kuttel; Investigation, Jason Hlozek and Sara Owen; Methodology, Neil Ravenscroft and Michelle M. Kuttel; Project administration, Michelle M. Kuttel; Resources, Neil Ravenscroft and Michelle M. Kuttel; Software, Michelle M. Kuttel; Supervision, Neil Ravenscroft and Michelle M. Kuttel; Writing – original draft, Jason Hlozek and Sara Owen; Writing – review & editing, Neil Ravenscroft and Michelle M. Kuttel.

Funding: This work was supported by the National Research Foundation of South Africa [Grant numbers 111704 to N.R., 103805 to M.M.K., 109643 to J.H.] and University of Cape Town PhD scholarship funding.

Acknowledgments: Computations were performed using facilities provided by the University of Cape Town's ICTS High Performance Computing team: http://hpc.uct.ac.za.

Conflicts of Interest: The authors declare no conflict of interest.

5.8 Supplementary information



Supplementary Figure S5.1: Block averaging analysis for the (a) end to end distance, r, and (b) radius of gyration for 6 RUs of the serogroup 3 and 5 O-Ags. The plot of blocked standard error (BSE) versus block size reaches a plateau for all the O-Ags of serogroup 3 and 5, indicating convergence in each simulation.

	1A_2B	1A_1D	2A_1D	2A_1F	3A_1F	1B_2C	1B_3C	1C_3D	2C_5B	1D_1F
Average distance[a]	2.3	3.9	2.3	2.3	2.7	3.7	2.3	2.2	2.7	2.8
r^6 average MD[a]	2.3	3.8	2.2	2.3	2.6	3.3	2.2	2.2	2.5	2.6
Theillet NMR[b]	2.2	3.7	2.3	2.3	2.4	3.3	2.2	2.3	2.7	2.6
Theillet MD[b]	2.3	3.5	2.4	2.3	2.4	3.3	2.2	2.3	2.7	2.6

Supplementary Table S5.2: Comparison between MD simulation and experimental data of distances between hydrogen atom pairs (columns) are in good agreement for serotype 3a. (a) The mean distance and corresponding r6 average from our 1.2 µs production runs, excluding 200 ns for equilibration. (b) Experimental NMR distances from NOE build-up curves (estimated 10% error) and distances calculated by full ensemble relaxation analysis of 60 ns MD simulations of a 3RU serotype 3a O-Ag (Theillet et. al., 2010).

	1A_1B	1A_2B	1A_4B	1A_1D	1A_2D	2A_1B	2A_1D	2A_2D	2A_5D	2A_1E	4A_1D	5A_1B	5A_2B	1B_2C	1B_3C	1B_4C	2B_4C	2B_1E	3B_1E	3B_5E
Average distance[a]	3.6	2.3	4.2	3.6	4.4	5	2.4	4.3	4	4.5	4.2	2.6	3.7	4	2.3	3.9	4.2	2.4	2.7	3
r^6 average[a]	3.2	2.2	4	3.1	4.3	4.9	2.3	4.3	3.9	3.9	4	2.4	3.4	3.6	2.3	3.7	4.1	2.3	2.6	2.7
ROE[b]	3.1	2.1	4.7		4.5		2	4.5	4.1	4.1	4.2	2.4	4.6	3.4	2.1	4.3	4.6	2.3	2.5	3
NOE[b]	3	2.1	4.6	3.5		4.7	2.1	4.5	3.8	4.1	4.3	2.6	4.3	2.9	2.1	4.4	4.8	2.4	2.2	
Theillet MD[c]	3.2	2.3	4.5	3.2		4.9	2.3	4.5	4.1	4	4.2	2.4	3.9	3.3	2.3	4	4.4	2.3	2.6	

Supplementary Table S3: Distances of pairs of hydrogen atoms (columns) from MD simulation and experimental data for serotype 5a. (a) The average distance and corresponding r6 average from our 2 µs production runs, excluding 200 ns for equilibration. (b) Experimental NMR distances (estimated 10% error) derived from NOE build-up curves and ROESY off-resonance measurements (Clement et. al., 2003). (c) Distances calculated by full ensemble relaxation analysis of 60 ns MD simulations of a 3RU O-Ag (Theillet et. al., 2010).

Chapter 6

Discussion and conclusions

This dissertation employed molecular dynamics simulations to investigate the polysaccharide antigens of prevalent disease-causing *N. meningitidis* serogroups and *S. flexneri* serotypes to promote a rational basis for vaccine strain selection. The computational modeling in this work provides detailed descriptions of the dynamic motion of glycan chains at the molecular scale and identifies the effects of substitution on backbone conformation in chemically similar glycan chains. This theoretical understanding of glycan antigen conformations aims to guide the design and composition of future glycoconjugate vaccines against these pathogens.

Our methodology characterizes glycan conformations with molecular dynamics using an incremental approach - beginning with brief simulations of short chains and building up to extended chains with long production runs. This approach produces initial glycan structures with low-energy conformations and, therefore, the simulation trajectory is likely to converge quickly and explore the most prevalent conformations of the native glycan chains in solution.

The computational methodology in each study made use of a set of metrics to characterize and compare the dynamic behavior of each glycan chain. The end-to-end distance is a simple metric that gives a broad overview of relative chain extension and flexibility. Clustering analysis then reveals specific detail of overall glycan chain behavior by grouping snapshots of the glycan chain over the simulation trajectory into

conformational families. Although the mechanisms of antigen cross-protection are not well understood, it seems reasonable to assume that similarity in chain conformation is required, considering the role of antigen-binding to cross-protective antibodies in the immune response. It is on this conformational basis that theoretical insight can be gleaned to guide vaccine composition and design. Finally, we investigate a rationale for the specific structural consequences of chemical alterations in a glycan RU through analysis of each glycosidic linkage. As the majority of glycan motion arises from transitions in the glycosidic bond orientations, scatter plots of the linkage torsion angles over the simulation trajectory can uncover subtle mechanistic details and atomic interactions that contribute to the differences observed in overall chain conformation. This specific detail can identify targets for synthetic chemical alterations of a native RU in the development of antigen mimics that maintain conformational similarity.

Each individual metric supplies partial insight into the flexibility of a glycan antigen, however, combined they provide a thorough theoretical characterization of the overall conformational behavior. The comparative methodology employed in this work – juxtaposing chemically similar glycan chains under identical simulation conditions - is powerful in highlighting relative conformational differences in preferred chain conformations and provides a rational basis for vaccine antigen selection - including in the absence of experimental data.

If limited or no experimental data is available, alternative statistical methods can be employed to validate simulation data and subsequent conformational findings. For example, to compensate for the lack of substantial experimental findings for *Shigella*

cross-protection in human studies, we undertook additional statistical analysis (block averaging analysis) to evaluate the extent of conformational sampling in each simulation and the sufficiency of the chosen equilibration time. Further, these statistical measures proved useful to inform decisions relating to simulation length - whether extension was required or if simulation convergence had occurred and analysis could therefore begin. Assessment of the statistical fitness of a data set is an important additional step to include in our systematic methodology for the robustness of future modeling studies, even when experimental data is available for corroboration.

It should be emphasized that combining all available data - both theoretical and experimental - is the most powerful approach to building a holistic understanding of carbohydrate conformation and antigen cross-protection. For instance, computational modeling alone cannot account for confounding immunological phenomena such as immunodominance, where the strength of an immune response varies for specific epitopes. However, insight from MD remains crucial to the understanding of glycan antigens as empirical findings specifically investigating serotype conformation or cross-protection tend to be uncommon and experimental data (e.g. X-ray crystallography and NMR) only provides limited descriptions of dynamic motion.

While static glycan X-ray crystal data can allow for construction of three-dimensional models, the insight is limited as it does not reveal the extent of possible dynamic motion or indicate structural features arising from constraints in the crystal lattice that might not be present in solution. However, this technique can provide important insights into the relative sizes of polysaccharide epitopes, which can inform future MD

simulations. Furthermore, bound antibody-antigen crystal structures, if available, can still be particularly useful in revealing important binding modes and corroborating unusual linkage conformations such as for the branching glucose of *S. flexneri* serotype 2a described in chapter 4. Unfortunately, glycan crystal structure data is rare due to challenges in the crystallization of carbohydrates.

Multiple NMR techniques are available that each provide partial insight into molecular conformation. NOE correlations, arising from the transferal of magnetization of close-contact spin-active nuclei in NMR experiments, are a common source of experimental corroboration for predicted structures. However, computational models of glycan conformations are still required to contextualize this experimental data as NOE values only arise over short atomic distances (< 5 Å), which typically provides insufficient correlations to identify a specific chain conformation.

Another conformational insight provided by NMR is the three-bond coupling (J^3) between spin-active nuclei, which correspond to an estimated torsion angle via a Karplus equation that is parameterized according to the identity of the atoms involved in bonding. While the coupling constant can be measured very accurately, the calculation of the corresponding torsion angle is ambiguous as the sinusoidal nature of a Karplus equation relates each coupling constant to multiple possible torsion angles. However, these coupling constants still allow for the exclusion of accessible regions on a phi/psi plot and can be compared to the predicted average from MD simulations for model validation.

A more recently developed technique, is to measure residual dipolar couplings (RDCs) for relatively rigid structures, which provides complementary conformational data to

NOEs and J^3 coupling. By disturbing the isotropic nature of a solution through the addition of a solute that aligns with the external NMR magnetic field, RDCs can be measured from NMR spectra. The RDCs encode the orientation of specific atom pairs relative to the external magnetic field. With an r^3 distance dependence, RDCs can also be measured over much longer distances than NOEs, which arise from an r^6 distance dependence. Provided sufficient sample is available for the multiple experiments required, RDCs have proven a powerful tool in protein research.[167]

However, conformational insights from NMR have inherent ambiguity as the experimental data can only represent the average molecular conformation in solution and can be misleading if the molecule is particularly flexible. A measurement arising from a relatively static molecular interaction is indistinguishable from a measurement representing the weighted average of a pair of atoms with significant molecular motion. For this reason, NMR data frequently yields measurements for use as restraint data in structural refinement but is limited in revealing dynamic motion without a computational model for context. Therefore, MD provides unique insight into dynamic molecular motion, while experimental methods tend to play a larger role in the testing and validation of the MD simulations.

The central insight from this computational work are descriptions of the range of preferred glycan antigen conformations and their relative occupancies (corresponding to the energetic stability) that rationalize and predict antigen cross-protection. MD can investigate further pertinent questions for vaccine development in collaboration with experiment. For instance, optimal chain length is a particularly important consideration for vaccines composed of synthetic glycans: a shorter chain length

affords less complex synthesis and improved yields but may not elicit an immune response that will recognize the longer chains of the native antigen. While the immune response is inherently an empirical finding, our incremental MD methodology is well suited to highlighting changes in glycan conformation as a function of length, as illustrated in the modeling of *N. meningitidis* glycan antigens in chapters 2 and 3. Likewise, molecular dynamics can provide complementary understanding of antigenicity findings by investigating the solvent accessibility of specific epitopes (e.g. O-acetylation in *N. meningitidis* and *S. flexneri*) as well as the conformations of antigen mimics, as presented in this body of work. By first investigating these phenomena computationally, laboratory time can be focused towards the most relevant areas for experimental investigation, optimizing laboratory resources.

A weakness of the MD methodology is the time intensive nature in computing simulation data sets of sufficient length, which depends on the level of detail required and the computational power available. This limitation can be mitigated by employing further 'coarse-grained' approximations to reduce computation time and by planning experimental studies to run in parallel with computational experiments. However, obtaining computational insight in a timely manner remains an important barrier to the more wide-spread adoption of molecular dynamics simulations. Nevertheless, molecular modeling continues to contribute key understanding of dynamic glycan conformations, both complimenting experimental investigation and as a standalone methodology.

Overall, the multiple case studies of glycan antigen conformations presented in this work describe multiple examples of key theoretical insights provided by MD

simulations that are typically inaccessible to direct empirical study. These investigations develop cross-protection rationale based on conformational analysis, identify mechanisms inducing conformational change from chemical alterations in a RU, discern dependence of conformational behavior on antigen chain-length and reveal the solvent accessibility of immunologically important epitopes. Specifically, our findings rationalize the cross-protection data available for *N. meningitidis* serogroups A and X and has a more predictive role for the expected cross-protection between *S. flexneri* O-Ags, resulting in recommendations for the design of vaccines against each pathogen. Ultimately, the significant explanatory power of the computational methodology employed in these studies promotes and builds a rational basis for glycan vaccine design instead of the empirical 'trial-and-error' approach of decades past.

The first pair of publications consider the antigens of *N. meningitidis* serogroup A and X with CPSs consisting of phosphodiester-linked monosaccharide RUs. The first study compared the conformations of oligosaccharide chains of meningococcal serogroup A and X, which differ in RU linkage position and N-acetyl orientation at C-2 (1->6 linked ManNAc and 1->4 linked GlcNAc for A and X, respectively). Serogroup X was found to adopt helical epitopes that were corroborated by key NMR NOESY correlations. By contrast, the five-bond α(1->6) linkage of serogroup A had significant flexibility with random-coil behavior, as exhibited in the wide array of conformational families.

Investigation of chemical substitutions to the serogroup A RU found notable shifts in conformation relative to the unsubstituted backbone. Firstly, O-acetylation on the native serogroup A chain causes a significant straightening effect, inducing rigidity in the highly flexible backbone. Secondly, a serogroup A mimic (with substitution of the

ring-oxygen for a methylene group) was found to maintain considerable chain flexibility, however, the chemical alteration induced significant localized conformational differences that inhibited S-bend conformations in the carba-MenA mimic that were present in the native MenA chain.

These noteworthy findings of conformational differences in *N. meningitidis* antigens relative to the serogroup A chain provides guidance for the composition of future multivalent meningococcal vaccines: 1) A serogroup A vaccine is recommended to possess O-3-acetylation. The native antigen is O-3-acetylated (70% - 95%), which we find causes a pronounced straightening effect in the serogroup A chain conformation. This supplies corroboration for the requirement of O-acetylation (> 61.5%) in serogroup A vaccines because a vaccine with the same dominant conformation as the native antigen is likely to provide optimal protection. 2) A carba-MenA mimic, despite increased chemical stability, is unlikely to provide strong cross-protection against serogroup A and is not recommended as a replacement serogroup A vaccine candidate due to the significant localized conformational differences. 3) Serogroup X is recommended to be included as a distinct component in the next generation of vaccines to ensure broad coverage against meningococcal infection. Despite both serogroup A and X consisting of phosphodiester-linked monosaccharide RUs, the lack of conformational similarity arising from differing linkage positions suggests that little or no cross-protection will occur from serogroup A vaccination. 4) Based on literature studies on 1 RU and our work on 2 RUs and 6 RUs, meningococcal serogroup A and X glycan vaccine components should consist of longer chains (approximately ≥ 6 RUs) as

the conformations of the shorter glycan chains are not found to be representative of the conformations of the longer native CPS chains.

The final two papers explore the potential for cross-protection between *S. flexneri* serotypes to guide current development efforts towards an envisioned broad-coverage multivalent *Shigella* vaccine with minimal valency. As each strain shares the same →2)-α-L-Rhap^{III}-(1→2)-α-L-Rhap^{II}-(1→3)-α-L-Rhap^{I}-(1→3)-β-D-GlcpNAc-(1→ backbone (residues A-B-C-D), cross-protection between strains is expected; however, only limited clinical cross-protection data is available to illustrate this. The third publication in this dissertation investigates the conformations of the most prevalent *Shigella* strains – the serogroup 2 O-Ags - as well as the unsubstituted serotype Y backbone. The investigation found the serotype Y backbone to be highly flexible, while the serogroup defining glucosylation of serotypes 2a and 2b significantly constrains backbone motion to C-curves and helices, respectively. Most significantly, the study discovered that an O-3-acetylated serotype 2a chain shares conformational features with both serotypes 2a and 2b, acting as a conformational hybrid between the two prevalent infectious serotypes. This suggests that an O-3-acetylated 2a vaccine could function as an optimized antigen in a first *Shigella* vaccine, allowing for broad serogroup 2 cross-protection from a single vaccine component.

The fourth and final publication expanded the study of *S. flexneri* O-antigens by considering the glycan RUs of serogroups 3 and 5, in addition to the previous serogroup 2 findings, within the context of a proposed broad-coverage vaccine containing serotypes 2a and 3a. This broad investigation of *Shigella* O-Ags found guiding heuristics for the effects on backbone conformation by substitution at varying

positions in the O-Ag backbone RU. First, substitution on rhamnose C, whether O-4-glucosylation (serotype 2a) or O-2-acetylation (serotype 3a), had the most significant effect on chain conformation - restricting the glycan chain in both cases to predominantly C-curves conformational families. Second, O-3-glucosylation on rhamnose A (serotypes 2a-3Ac, 2b, 3a and 5b) restricts the chain to predominantly helical conformations through steric interaction with the adjacent glucose-N-acetyl residue. Thirdly, O-3-glucosylation on rhamnose B (serogroup 5) has only a minor effect on O-Ag conformation.

These studies on the conformations of immunologically important *Shigella* O-antigens provide predictive insight into potential cross-reactivity from an anticipated multivalent *Shigella* vaccine (containing serotypes 2a, 3a, 6 and *S. sonnei*) as well as provide recommendations for the composition of such a vaccine, as follows: 1) A serotype 2a component should include O-3-acetylation on rhamnose A – in addition to direct coverage of the dominant strain causing *S. flexneri* infection (2a), the helical conformation would likely provide increased serotype 2b cross-protection and potential broader serogroup coverage through shared helical conformations in serogroups 3 and 5. 2) Serotype 3a is a recommended second component for a multivalent vaccine. Given the prevalence of serogroup 3 infection, our modeling indicates that serotype 3a could offer protection against the immunologically important 3b and may further enhance coverage through shared helical conformations that are induced by glucosylation on rhamnose A. 3) A serogroup 5 vaccine component is likely unnecessary given the low prevalence of infection and possible coverage from the serotype 2a-3Ac and 3a components.

The recommended vaccine components, serotypes 2a (with O-acetylation) and 3a, provide direct coverage against highly prevalent serotypes and exhibit the range of epitopes found on the remaining *S. flexneri* serotypes (except for serotype 6). As cross-protection against non-vaccine serotypes is supported by our conformational analysis of serogroups Y, 2, 3, and 5, which display shared conformations and similar conformational trends for given substitutions; a multivalent vaccine offering broad-coverage against *Shigella* infection is found to be promising.

Prominent topics arising as future work from this dissertation include:

For *Shigella*, the investigation of the conformations of *S. flexneri* O-Ags can be further extended from our study of serogroups 2, 3 and 5 (Fig. 6.1a) to the remaining prevalent strains – serogroup 1 and serotype 6. Serogroup 1 (Fig. 6.1b) is defined by glucosylation at O-4 of GlcNAc D, which could expand our set of heuristics to include the effects due to conformational shifts in the C-D linkage. However, the shared substitutions at rhamnose A and C, affords insight from our established heuristics. According to our first heuristic, O-acetylation on rhamnose C (serotype 1b) is predicted to restrict the chain to predominantly C-curve conformations. Substitution on O-3 of rhamnose A (O-acetylation in 1a and 1b as well as glucosylation for 1d) are predicted to have helical conformations as in serotypes 2a-3Ac, 2b, 3a and 5b. Serotype 6 (Fig. 6.1c) is also projected to have a helical conformation with O-3-acetylation on rhamnose A but to ascertain the effect of the alterations in backbone structure (residues C and D) will require further inquiry.

Modeling of the conformations of these serotypes can further refine our set of heuristics as well as analyze the expected cross-protection from a *Shigella* vaccine

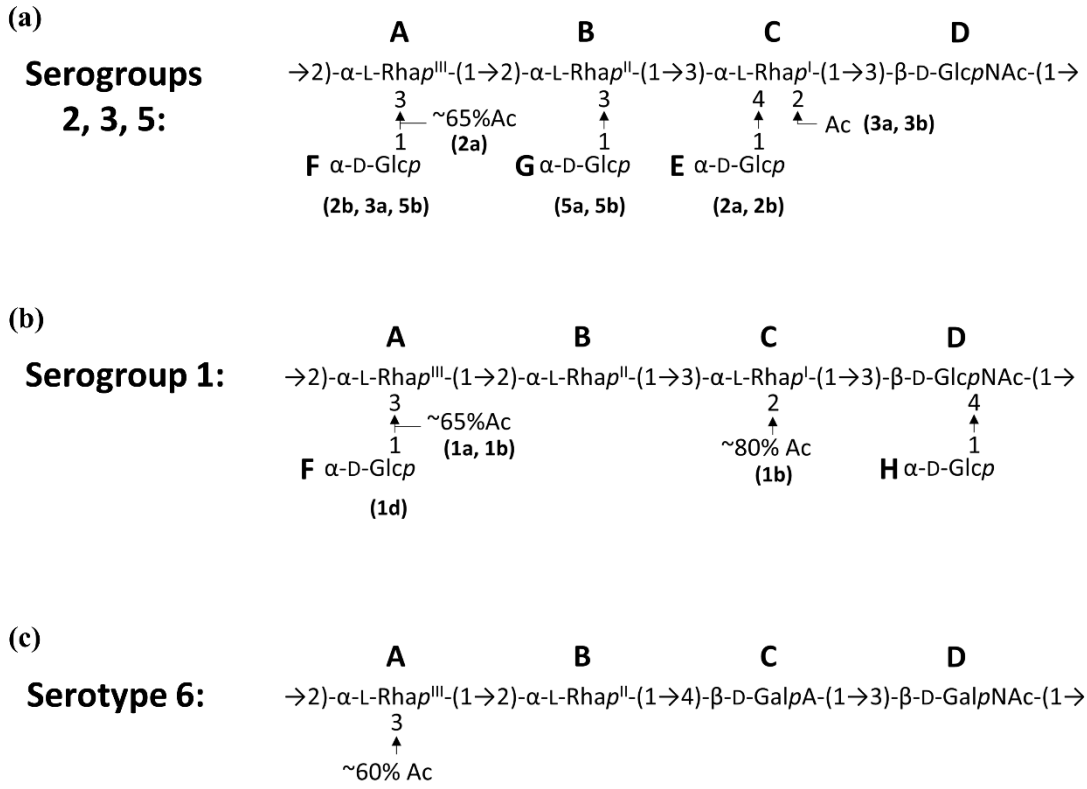


Figure 6.1: Line structures of serogroups 2, 3 and 5 (a) studied in this work as well as the prevalent serogroup 1 (b) and serotype 6 (c) strains suggested for future study.

consisting of an O-3-acetylated serotype 2a and serotype 3a (with serotype 6 and *S. sonnei*) against the remaining non-vaccine O-Ags.

For *N. meningitidis* serogroup A (Fig. 6.2a), while the second publication found the carba-MenA analogue (Fig. 6.2b) to be unlikely to elicit strong cross-protection against serogroup A infection, further study is needed to investigate the detected bactericidal response for a Carba-MenA chain of moderate length and determine whether length or conformation is the critical factor determining immunogenicity. As a chemically-stabilized MenA vaccine antigen is still required, MD simulations could also investigate whether a phosphono-meningococcal serogroup (Fig. 6.2c), with methylene

substitution in the phosphodiester linkage, could exhibit greater conformational similarity with the native serogroup A RU than the Carba-MenA mimic.

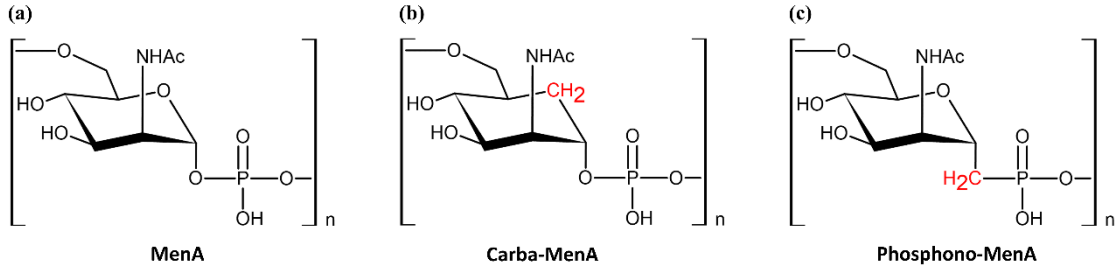


Figure 6.2: Monosaccharide RUs of the (a) native serogroup A, (b) carba-MenA and (c) phosphono-MenA. The substitution of an oxygen atom for a methylene group is marked in red for the Carba-MenA and Phosphono-MenA RUs.

In addition to these studies, further experimental evidence can be sought to corroborate the theoretical aspects of molecular modeling. As computational power increases, MD simulations can investigate larger systems in ever-greater detail and can play a more widespread role in guiding vaccine development and rational antigen selection.

References

1. Hlozek, J.; Kuttel, M. M.; Ravenscroft, N., Conformations of Neisseria meningitidis serogroup A and X polysaccharides: The effects of chain length and O-acetylation. *Carbohydr. Res.* **2018,** *465*, 44-51.

2. Hlozek, J.; Ravenscroft, N.; Kuttel, M. M., Modeling the conformations of Neisseria meningitidis serogroup a CPS and a carba-analogue: Implications for vaccine development. *Carbohydr. Res.* **2019,** *486*, 107838.

3. Hlozek, J.; Ravenscroft, N.; Kuttel, M. M., Effects of Glucosylation and O-Acetylation on the Conformation of Shigella flexneri Serogroup 2 O-Antigen Vaccine Targets. *J. Phys. Chem. B* **2020,** *124* (14), 2806-2814.

4. Ravenscroft, N.; Omar, A.; Hlozek, J.; Edmonds-Smith, C.; Follador, R.; Serventi, F.; Lipowsky, G.; Kuttel, M. M.; Cescutti, P.; Faridmoayer, A., Genetic and structural elucidation of capsular polysaccharides from Streptococcus pneumoniae serotype 23A and 23B, and comparison to serotype 23F. *Carbohydr. Res.* **2017,** *450*, 19-29.

5. Rémy, V.; Zöllner, Y.; Heckmann, U., Vaccination: the cornerstone of an efficient healthcare system. *J. Mark. Access Health Policy* **2015,** *3* (1), 27041.

6. Ehreth, J., The global value of vaccination. *Vaccine* **2003,** *21* (7-8), 596-600.

7. W.H.O., Global vaccine action plan 2011-2020. **2013**.

8. Andre, F. E.; Booy, R.; Bock, H. L.; Clemens, J.; Datta, S. K.; John, T. J.; Lee, B. W.; Lolekha, S.; Peltola, H.; Ruff, T., Vaccination greatly reduces disease, disability, death and inequity worldwide. *Bulletin of the W.H.O.* **2008,** *86*, 140-146.

9. Plotkin, S., History of vaccination. *PNAS* **2014,** *111* (34), 12283-12287.

10. Black, R. E.; Cousens, S.; Johnson, H. L.; Lawn, J. E.; Rudan, I.; Bassani, D. G.; Jha, P.; Campbell, H.; Walker, C. F.; Cibulskis, R., Global, regional, and national causes of child mortality in 2008: a systematic analysis. *Lancet* **2010,** *375* (9730), 1969-1987.

11. Odone, A.; Ferrari, A.; Spagnoli, F.; Visciarelli, S.; Shefer, A.; Pasquarella, C.; Signorelli, C., Effectiveness of interventions that apply new media to improve vaccine uptake and vaccine coverage: a systematic review. *Hum. Vaccin. Immunother.* **2015,** *11* (1), 72-82.

12. Lahariya, C., Vaccine epidemiology: a review. *J. Family Med. Prim. Care* **2016,** *5* (1), 7.

13. Turner, H. C.; Thwaites, G. E.; Clapham, H. E., Vaccine-preventable diseases in lower-middle-income countries. *Lancet Infect. Dis.* **2018,** *18* (9), 937-939.

14. Feldstein, L. R.; Mariat, S.; Gacic-Dobo, M.; Diallo, M. S.; Conklin, L. M.; Wallace, A. S., Global routine vaccination coverage, 2016. *MMWR. Morbidity and mortality weekly report* **2017,** *66* (45), 1252.

15. W.H.O., Global Health Estimates 2016: Deaths by Cause, Age, Sex, by Country and by Region, 2000-2016. Geneva: WHO, 2018. 2019.

16. Tadesse, B. T.; Ashley, E. A.; Ongarello, S.; Havumaki, J.; Wijegoonewardena, M.; González, I. J.; Dittrich, S., Antimicrobial resistance in Africa: a systematic review. *BMC Infect. Dis.* **2017,** *17* (1), 616.

17. Bernabe, K. J.; Langendorf, C.; Ford, N.; Ronat, J.-B.; Murphy, R. A., Antimicrobial resistance in West Africa: a systematic review and meta-analysis. *Int. J. Antimicrob. Agents* **2017,** *50* (5), 629-639.

18. Reckseidler-Zenteno, S. L., Capsular polysaccharides produced by the bacterial pathogen Burkholderia pseudomallei. In *The Complex World of Polysaccharides*, InTech: 2012.

19. Hennessey Jr, J. P.; Costantino, P.; Talaga, P.; Beurret, M.; Ravenscroft, N.; Alderson, M. R.; Zablackis, E.; Prasad, A. K.; Frasch, C., Lessons learned and future challenges in the design and manufacture of glycoconjugate vaccines. In *Carbohydrate-Based Vaccines: From Concept to Clinic*, ACS Publications: 2018; pp 323-385.

20. Costantino, P.; Rappuoli, R.; Berti, F., The design of semi-synthetic and synthetic glycoconjugate vaccines. *Expert Opin. Drug. Discov.* **2011,** *6* (10), 1045-1066.

21. Micoli, F.; Costantino, P.; Adamo, R., Potential targets for next generation antimicrobial glycoconjugate vaccines. *FEMS Microbiol. Rev.* **2018,** *42* (3), 388-423.

22. Kuttel, M. M.; Ravenscroft, N., The Role of Molecular Modeling in Predicting Carbohydrate Antigen Conformation and Understanding Vaccine Immunogenicity. In *Carbohydrate-Based Vaccines: From Concept to Clinic*, ACS Publications: 2018; pp 139-173.

23. Kaplan, S. L.; Barson, W. J.; Lin, P. L.; Stovall, S. H.; Bradley, J. S.; Tan, T. Q.; Hoffman, J. A.; Givner, L. B.; Mason, E. O., Serotype 19A is the most common serotype causing invasive pneumococcal infections in children. *Pediatrics* **2010,** *125* (3), 429-436.

24. Lee, H.; Nahm, M. H.; Burton, R.; Kim, K.-H., Immune response in infants to the heptavalent pneumococcal conjugate vaccine against vaccine-related serotypes 6A and 19A. *Clin. Vaccine Immunol.* **2009,** *16* (3), 376-381.

25. Kuttel, M. M.; Jackson, G. E.; Mafata, M.; Ravenscroft, N., Capsular polysaccharide conformations in pneumococcal serotypes 19F and 19A. *Carbohydr. Res.* **2015,** *406*, 27-33.

26. Griffiss, J.; Brandt, B.; Altieri, P.; Pier, G.; Berman, S., Safety and immunogenicity of group Y and group W135 meningococcal capsular polysaccharide vaccines in adults. *Infect. Immun.* **1981,** *34* (3), 725-732.

27. Kuttel, M. M.; Timol, Z.; Ravenscroft, N., Cross-protection in Neisseria meningitidis serogroups Y and W polysaccharides: A comparative conformational analysis. *Carbohydr. Res.* **2017,** *446*, 40-47.

28. Finco, O.; Rappuoli, R., Designing vaccines for the twenty-first century society. *Front. Immunol.* **2014,** *5*, 12.

29. Rueckert, C.; Guzmán, C. A., Vaccines: from empirical development to rational design. *PLoS Pathog.* **2012,** *8* (11).

30. Guvench, O.; Mallajosyula, S. S.; Raman, E. P.; Hatcher, E.; Vanommeslaeghe, K.; Foster, T. J.; Jamison, F. W.; MacKerell Jr, A. D., CHARMM additive all-atom force field for

carbohydrate derivatives and its utility in polysaccharide and carbohydrate–protein modeling. *J. Chem. Theory Comput.* **2011,** *7* (10), 3162-3180.

31. Guvench, O.; Greene, S. N.; Kamath, G.; Brady, J. W.; Venable, R. M.; Pastor, R. W.; Mackerell Jr, A. D., Additive empirical force field for hexopyranose monosaccharides. *J. Comput. Chem.* **2008,** *29* (15), 2543-2564.

32. Guvench, O.; Hatcher, E.; Venable, R. M.; Pastor, R. W.; MacKerell Jr, A. D., CHARMM additive all-atom force field for glycosidic linkages between hexopyranoses. *J. Chem. Theory Comput.* **2009,** *5* (9), 2353-2370.

33. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781-1802.

34. Lindahl, E., Molecular dynamics simulations. In *Molecular modeling of proteins*, Springer: 2015; pp 3-26.

35. Grossfield, A.; Zuckerman, D. M., Quantifying uncertainty and sampling quality in biomolecular simulations. *Annu. Rep. Comput. Chem.* **2009,** *5*, 23-48.

36. Laio, A.; Parrinello, M., Escaping free-energy minima. *PNAS* **2002,** *99* (20), 12562-12566.

37. Yazdankhah, S. P.; Caugant, D. A., Neisseria meningitidis: an overview of the carriage state. *J. Med. Microbiol.* **2004,** *53* (9), 821-832.

38. Caugant, D. A.; Maiden, M. C., Meningococcal carriage and disease—population biology and evolution. *Vaccine* **2009,** *27*, B64-B70.

39. Flexner, S.; Jobling, J. W., An analysis of four hundred cases of epidemic meningitis treated with the anti-meningitis serum. *J. Exp. Med.* **1908,** *10* (5), 690-733.

40. Trotter, C. L.; Greenwood, B. M., Meningococcal carriage in the African meningitis belt. *Lancet Infect. Dis.* **2007,** *7* (12), 797-803.

41. Molesworth, A. M.; Thomson, M. C.; Connor, S. J.; Cresswell, M. P.; Morse, A. P.; Shears, P.; Hart, C. A.; Cuevas, L. E., Where is the meningitis belt? Defining an area at risk of epidemic meningitis in Africa. *Trans R Soc Trop Med Hyg.* **2002,** *96* (3), 242-249.

42. Borrow, R.; Alarcón, P.; Carlos, J.; Caugant, D. A.; Christensen, H.; Debbag, R.; De Wals, P.; Echániz-Aviles, G.; Findlow, J.; Head, C., The Global Meningococcal Initiative: global epidemiology, the impact of vaccines on meningococcal disease and the importance of herd protection. *Expert Rev. Vaccines* **2017,** *16* (4), 313-328.

43. Goldschneider, I.; Lepow, M. L.; Gotschlich, E. C.; Mauck, F. T.; Bachl, F.; Randolph, M., Immunogenicity of group A and group C meningococcal polysaccharides in human infants. *J. Infect. Dis.* **1973,** *128* (6), 769-776.

44. Wong, K.; Barrera, O.; Sutton, A.; May, J.; Hochstein, D.; Robbins, J.; Robbins, J.; Parkman, P.; Seligmann, E., Standardization and control of meningococcal vaccines, group A and group C polysaccharides. *J. Biol. Stand.* **1977,** *5* (3), 197-215.

45. Pollard, A. J.; Feavers, I.; Cohn, A., Prevention of meningococcal disease through vaccination. In *Handbook of Meningococcal Disease Management*, Springer: 2016; pp 91-103.

46. Frasch, C.; Preziosi, M.-P.; LaForce, F. M., Development of a group A meningococcal conjugate vaccine, MenAfriVacTM. *Hum. Vaccin. Immunother.* **2012,** *8* (6), 715-724.

47. LaForce, F. M.; Ravenscroft, N.; Djingarey, M.; Viviani, S., Epidemic meningitis due to Group A *Neisseria meningitidis* in the African meningitis belt: a persistent problem with an imminent solution. *Vaccine* **2009,** *27*, B13-B19.

48. Novak, R. T.; Kambou, J. L.; Diomandé, F. V.; Tarbangdo, T. F.; Ouédraogo-Traoré, R.; Sangaré, L.; Lingani, C.; Martin, S. W.; Hatcher, C.; Mayer, L. W., Serogroup A meningococcal conjugate vaccination in Burkina Faso: analysis of national surveillance data. *Lancet Infect. Dis.* **2012,** *12* (10), 757-764.

49. Kristiansen, P. A.; Diomandé, F.; Ba, A. K.; Sanou, I.; Ouédraogo, A. S.; Ouédraogo, R.; Sangaré, L.; Kandolo, D.; Aké, F.; Saga, I. M., Impact of the serogroup A meningococcal

conjugate vaccine, MenAfriVac, on carriage and herd immunity. *Clin. Infect. Dis.* **2012,** *56* (3), 354-363.

50. Kristiansen, P. A.; Ba, A. K.; Ouédraogo, A.-S.; Sanou, I.; Ouédraogo, R.; Sangaré, L.; Diomandé, F.; Kandolo, D.; Saga, I. M.; Misegades, L., Persistent low carriage of serogroup A *Neisseria meningitidis* two years after mass vaccination with the meningococcal conjugate vaccine, MenAfriVac. *BMC Infect. Dis.* **2014,** *14* (1), 663.

51. Xie, O.; Pollard, A. J.; Mueller, J. E.; Norheim, G., Emergence of serogroup X meningococcal disease in Africa: need for a vaccine. *Vaccine* **2013,** *31* (27), 2852-2861.

52. Soeters, H. M.; Yameogo, I.; Sawadogo, G.; Aké, F.; Lingani, C.; Wang, X.; Bita, A.; Fall, A.; Sangaré, L.; Ouédraogo-Traoré, R., Bacterial meningitis epidemiology and return of Neisseria meningitidis serogroup A cases in Burkina Faso in the five years following MenAfriVac mass vaccination campaign. *PLoS ONE* **2017,** *12* (11), e0187466.

53. La Force, M. In *Progress on MenACWYX vaccine: phase 1 trial results.* , Proceedings of MRF Conference, London, November 14-15; London, 2017.

54. Berry, D. S.; Lynn, F.; Lee, C.-H.; Frasch, C. E.; Bash, M. C., Effect of O acetylation of Neisseria meningitidis serogroup A capsular polysaccharide on development of functional immune responses. *Infect. Immun.* **2002,** *70* (7), 3707-3713.

55. W.H.O., Recommendations to assure the quality, safety and efficacy of group A meningococcal conjugate vaccines. *World Health Organization, Geneva* **2006**.

56. Costantino, P.; Norelli, F.; Giannozzi, A.; D'Ascenzi, S.; Bartoloni, A.; Kaur, S.; Tang, D.; Seid, R.; Viti, S.; Paffetti, R., Size fractionation of bacterial capsular polysaccharides for their use in conjugate vaccines. *Vaccine* **1999,** *17* (9-10), 1251-1263.

57. Morelli, L.; Cancogni, D.; Tontini, M.; Nilo, A.; Filippini, S.; Costantino, P.; Romano, M. R.; Berti, F.; Adamo, R.; Lay, L., Synthesis and immunological evaluation of protein conjugates of Neisseria meningitidis X capsular polysaccharide fragments. *Beilstein J. Org. Chem.* **2014,** *10*, 2367–2376.

58. Gao, Q.; Tontini, M.; Brogioni, G.; Nilo, A.; Filippini, S.; Harfouche, C.; Polito, L.; Romano, M. R.; Costantino, P.; Berti, F., Immunoactivity of protein conjugates of carba analogues from Neisseria meningitidis a capsular polysaccharide. *ACS Chem. Biol.* **2013,** *8* (11), 2561-2567.

59. Harale, K. R.; Rout, J. K.; Chhikara, M. K.; Gill, D. S.; Misra, A. K., Synthesis and immunochemical evaluation of a novel *Neisseria meningitidis* serogroup A tetrasaccharide and its conjugate. *Org. Chem. Front.* **2017,** *4* (12), 2348-2357.

60. Fiebig, T.; Romano, M. R.; Oldrini, D.; Adamo, R.; Tontini, M.; Brogioni, B.; Santini, L.; Berger, M.; Costantino, P.; Berti, F., An efficient cell free enzyme-based total synthesis of a meningococcal vaccine candidate. *NPJ Vaccines* **2016,** *1*, 16017.

61. Nahm, M. H.; Lin, J.; Finkelstein, J. A.; Pelton, S. I., Increase in the prevalence of the newly discovered pneumococcal serotype 6C in the nasopharynx after introduction of pneumococcal conjugate vaccine. *J. Infect. Dis.* **2009,** *199* (3), 320-325.

62. Grant, L. R.; O'Brien, S. E.; Burbidge, P.; Haston, M.; Zancolli, M.; Cowell, L.; Johnson, M.; Weatherholtz, R. C.; Reid, R.; Santosham, M., Comparative immunogenicity of 7 and 13-valent pneumococcal conjugate vaccines and the development of functional antibodies to cross-reactive serotypes. *PloS one* **2013,** *8* (9), e74906.

63. Widmalm, G., A perspective on the primary and three-dimensional structures of carbohydrates. *Carbohydr. Res.* **2013,** *378*, 123-132.

64. Lütteke, T., Analysis and validation of carbohydrate three-dimensional structures. *Acta Crystallogr. D* **2009,** *65* (2), 156-168.

65. Kuttel, M.; Gordon, M.; Ravenscroft, N., Comparative simulation of pneumococcal serogroup 19 polysaccharide repeating units with two carbohydrate force fields. *Carbohydr. Res.* **2014,** *390*, 20-27.

66. Seppälä, I.; Mäkelä, O., Antigenicity of dextran-protein conjugates in mice. Effect of molecular weight of the carbohydrate and comparison of two modes of coupling. *J. Immunol.* **1989,** *143* (4), 1259-1264.

67. Paoletti, L.; Kasper, D.; Michon, F.; DiFabio, J.; Jennings, H.; Tosteson, T.; Wessels, M., Effects of chain length on the immunogenicity in rabbits of group B Streptococcus type III oligosaccharide-tetanus toxoid conjugates. *J. Clin. Invest.* **1992,** *89* (1), 203-209.

68. Bundle, D. R.; Smith, I. C.; Jennings, H. J., Determination of the Structure and Conformation of Bacterial Polysaccharides by Carbon 13 Nuclear Magnetic Resonance STUDIES ON THE GROUP-SPECIFIC ANTIGENS OF NEISSERIA MENINGITIDIS SEROGROUPS A AND X. *J. Biol. Chem.* **1974,** *249* (7), 2275-2281.

69. Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K., Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.* **2007,** *28* (16), 2618-2640.

70. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

71. Kuttel, M. M.; Ståhle, J.; Widmalm, G., CarbBuilder: Software for building molecular models of complex oligo-and polysaccharide structures. *J. Comput. Chem.* **2016,** *37* (22), 2098-2105.

72. Kuttel, M.; Mao, Y.; Widmalm, G.; Lundborg, M. In *CarbBuilder: An adjustable tool for building 3D molecular structures of carbohydrates for molecular simulation*, Proceedings of 7th IEEE International Conference on e-Science Stockholm, Sweden, Stockholm, Sweden, 2011; pp 395-402.

73. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graph.* **1996,** *14* (1), 33-38.

74. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994,** *101* (5).

75. Feller, S. E.; Zhang, Y.; Pastor, R. W.; Brooks, B. R., Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* **1995,** *103* (11), 4613-4621.

76. Darden, T.; York, D.; Pedersen, L., Particle mesh Ewald: An N· log (N) method for Ewald sums in large systems. *J. Chem. Phys.* **1993,** *98* (12), 10089-10092.

77. Cross, S.; Kuttel, M. M.; Stone, J. E.; Gain, J. E., Visualisation of cyclic and multi-branched molecules with VMD. *J. Mol. Graph. Model.* **2009,** *28* (2), 131-139.

78. Heyer, L. J.; Kruglyak, S.; Yooseph, S., Exploring expression data: identification and analysis of coexpressed genes. *Genome Res.* **1999,** *9* (11), 1106-1115.

79. Lankhorst, P. P.; Haasnoot, C. A.; Erkelens, C.; Altona, C., Carbon-13 NMR in conformational analysis of nucleic acid fragments 2. A reparametrization of the Karplus equation for vicinal NMR coupling constants in CCOP and HCOP fragments. *J. Biomol. Struct. Dyn.* **1984,** *1* (6), 1387-1405.

80. Tan, L. K.; Carlone, G. M.; Borrow, R., Advances in the development of vaccines against Neisseria meningitidis. *N. Engl. J. Med.* **2010,** *362* (16), 1511-1520.

81. Berti, F.; Romano, M.; Micoli, F.; Pinto, V.; Cappelletti, E.; Gavini, M.; Proietti, D.; Pluschke, G.; MacLennan, C.; Costantino, P., Relative stability of meningococcal serogroup A and X polysaccharides. *Vaccine* **2012,** *30* (45), 6409-6415.

82. Slättegård, R.; Teodorovic, P.; Kinfe, H. H.; Ravenscroft, N.; Gammon, D. W.; Oscarson, S., Synthesis of structures corresponding to the capsular polysaccharide of Neisseria meningitidis group A. *Org. Biomol. Chem.* **2005,** *3* (20), 3782-3787.

83. Torres-Sanchez, M. I.; Zaccaria, C.; Buzzi, B.; Miglio, G.; Lombardi, G.; Polito, L.; Russo, G.; Lay, L., Synthesis and biological evaluation of phosphono analogues of capsular polysaccharide fragments from *Neisseria meningitidis* A. . *Chem. Eur. J.* **2007,** *13*, 6623–6635.

84. Gao, Q.; Zaccaria, C.; Tontini, M.; Poletti, L.; Costantino, P.; Lay, L., Synthesis and preliminary biological evaluation of carba analogues from Neisseria meningitidis A capsular polysaccharide. *Org. Biomol. Chem.* **2012,** *10* (33), 6673-6681.

85. Fallarini, S.; Buzzi, B.; Giovarruscio, S.; Polito, L.; Brogioni, G.; Tontini, M.; Berti, F.; Adamo, R.; Lay, L.; Lombardi, G., A Synthetic Disaccharide Analogue from Neisseria meningitidis A capsular polysaccharide stimulates immune cell responses and induces immunoglobulin G (IgG) production in mice when protein-conjugated. *ACS Infect. Dis.* **2015,** *1* (10), 487-496.

86. Vangala, M.; Hotha, S., Synthesis of 1, 2, 3-triazolyl analog of *Neisseria meningitidis* A capsular polysaccharide. *J. Carbohydr. Chem.* **2018,** *37* (7-8), 393-413.

87. Lay, L., Design and synthesis of Neisseria meningitides A carba analogues as stable antigens for antibacterial vaccines. In *19th European Carbohydrate Symposium*, Barcelona, Spain, 2017.

88. Jo, S.; Myatt, D.; Qi, Y.; Doutch, J.; Clifton, L. A.; Im, W.; Widmalm, G., Multiple Conformational States Contribute to the 3D Structure of a Glucan Decasaccharide: A Combined SAXS and MD Simulation Study. *J. Phys. Chem. B* **2018,** *122*, 1169-1175.

89. Hitri, K.; Kuttel, M.; Benedetto, G. D.; Lockyer, K.; Gao, F.; Hansal, P.; Rudd, T. R.; Beamish, E.; Rijpkema, S.; Ravenscroft, N.; Bolgiano, B., O-acetylation of typhoid capsular polysaccharide confers polysaccharide rigidity and immunodominance by masking additional epitopes. *Vaccine* **2019,** *37*, 3866-3875.

90. Aytenfisu, A. H.; Simon, R.; Jr., A. D. M., Impact of branching on the conformational heterogeneity of the lipopolysaccharide from *Klebsiella pneumoniae*: Implications for vaccine design. *Carbohydr. Res.* **2019,** *475*, 39-47.

91. Frank, M.; Collins, P. M.; Peak, I. R.; Grice, I. D.; Wilson, J. C., An Unusual Carbohydrate Conformation is Evident in *Moraxella catarrhalis* Oligosaccharides. *Molecules* **2015,** *20* (8), 14234-14253.

92. Toma, L.; Legnani, L.; Rencurosi, A.; Poletti, L.; Lay, L.; Russo, G., Modeling of synthetic phosphono and carba analogues of N-acetyl-α-D-mannosamine 1-phosphate, the repeating unit of the capsular polysaccharide from *Neisseria meningitidis* serovar A. *Org. Biomol. Chem.* **2009,** *7* (18), 3734-3740.

93. Calloni, I.; Unione, L.; Jiménez-Osés, G.; Corzana, F.; Del Bino, L.; Corrado, A.; Pitirollo, O.; Colombo, C.; Lay, L.; Adamo, R., The conformation of the mannopyranosyl phosphate repeating unit of the capsular polysaccharide of *Neisseria meningitidis* serogroup A and its carba-mimetic. *Eur. J. Org. Chem.* **2018,** *2018* (33), 4548-4555.

94. Adamo, R. In *Glycoconjugate vaccines: present and future*, 2019 Annual Meeting of the Society for Glycobiology, Milan, Italy, August 2019; Milan, Italy, 2019.

95. Kotloff, K. L.; Nataro, J. P.; Blackwelder, W. C.; Nasrin, D.; Farag, T. H.; Panchalingam, S.; Wu, Y.; Sow, S. O.; Sur, D.; Breiman, R. F., Burden and aetiology of diarrhoeal disease in infants and young children in developing countries (the Global Enteric Multicenter Study, GEMS): a prospective, case-control study. *Lancet* **2013,** *382* (9888), 209-222.

96. Khalil, I. A.; Troeger, C.; Blacker, B. F.; Rao, P. C.; Brown, A.; Atherly, D. E.; Brewer, T. G.; Engmann, C. M.; Houpt, E. R.; Kang, G., Morbidity and mortality due to shigella and enterotoxigenic Escherichia coli diarrhoea: the Global Burden of Disease Study 1990–2016. *Lancet Infect. Dis.* **2018,** *18* (11), 1229-1240.

97. Liu, L.; Johnson, H. L.; Cousens, S.; Perin, J.; Scott, S.; Lawn, J. E.; Rudan, I.; Campbell, H.; Cibulskis, R.; Li, M., Global, regional, and national causes of child mortality: an updated systematic analysis for 2010 with time trends since 2000. *Lancet* **2012,** *379* (9832), 2151-2161.

98. Livio, S.; Strockbine, N. A.; Panchalingam, S.; Tennant, S. M.; Barry, E. M.; Marohn, M. E.; Antonio, M.; Hossain, A.; Mandomando, I.; Ochieng, J. B., Shigella isolates from the global enteric multicenter study inform vaccine development. *Clin. Infect. Dis.* **2014,** *59* (7), 933-941.

99. Knirel, Y.; Sun, Q.; Senchenkova, S.; Perepelov, A.; Shashkov, A.; Xu, J., O-Antigen modifications providing antigenic diversity of Shigella flexneri and underlying genetic mechanisms. *Biochemistry-Moscow* **2015,** *80* (7), 901-914.

100. Perepelov, A. V.; Shekht, M. E.; Liu, B.; Shevelev, S. D.; Ledov, V. A.; Senchenkova, S. y. N.; L'vov, V. L.; Shashkov, A. S.; Feng, L.; Aparin, P. G., Shigella flexneri O-antigens revisited: final elucidation of the O-acetylation profiles and a survey of the O-antigen structure diversity. *FEMS Immunol. Med. Microbiol.* **2012,** *66* (2), 201-210.

101. Levine, M. M.; Kotloff, K. L.; Barry, E. M.; Pasetti, M. F.; Sztein, M. B., Clinical trials of Shigella vaccines: two steps forward and one step back on a long, hard road. *Nat. Rev. Microbiol.* **2007,** *5* (7), 540.

102. Sun, Q.; Knirel, Y. A.; Lan, R.; Wang, J.; Sof'ya, N. S.; Jin, D.; Shashkov, A. S.; Xia, S.; Perepelov, A. V.; Chen, Q., A novel plasmid-encoded serotype conversion mechanism through addition of phosphoethanolamine to the O-antigen of Shigella flexneri. *PloS one* **2012,** *7* (9), e46095.

103. Ye, C.; Lan, R.; Xia, S.; Zhang, J.; Sun, Q.; Zhang, S.; Jing, H.; Wang, L.; Li, Z.; Zhou, Z., Emergence of a new multidrug-resistant serotype X variant in an epidemic clone of Shigella flexneri. *J. Clin. Microbiol.* **2010,** *48* (2), 419-426.

104. Van De Verg, L. L.; Bendiuk, N. O.; Kotloff, K.; Marsh, M. M.; Ruckert, J. L.; Puryear, J. L.; Taylor, D. N.; Hartman, A. B., Cross-reactivity of Shigella flexneri serotype 2a O antigen antibodies following immunization or infection. *Vaccine* **1996,** *14* (11), 1062-1068.

105. Mani, S.; Wierzba, T.; Walker, R. I., Status of vaccine research and development for Shigella. *Vaccine* **2016,** *34* (26), 2887-2894.

106. Noriega, F. R.; Liao, F. M.; Maneval, D. R.; Ren, S.; Formal, S. B.; Levine, M. M., Strategy for cross-protection among Shigella flexneri serotypes. *Infect. Immun.* **1999,** *67* (2), 782-788.

107. Kotloff, K. L.; Riddle, M. S.; Platts-Mills, J. A.; Pavlinac, P.; Zaidi, A. K., Shigellosis. *Lancet* **2018,** *391* (10122), 801-812.

108. Varki, A.; Freeze, H. H.; Manzi, A. E., Overview of glycoconjugate analysis. *Curr. Protoc. Protein Sci.* **2009,** *57* (1), 12.1. 1-12.1. 10.

109. Phalipon, A.; Tanguy, M.; Grandjean, C.; Guerreiro, C.; Bélot, F.; Cohen, D.; Sansonetti, P. J.; Mulard, L. A., A synthetic carbohydrate-protein conjugate vaccine candidate against Shigella flexneri 2a infection. *J. Immunol.* **2009,** *182* (4), 2241-2247.

110. Riddle, M. S.; Kaminski, R. W.; Di Paolo, C.; Porter, C. K.; Gutierrez, R. L.; Clarkson, K. A.; Weerts, H. E.; Duplessis, C.; Castellano, A.; Alaimo, C., Safety and immunogenicity of a candidate bioconjugate vaccine against Shigella flexneri 2a administered to healthy adults: a single-blind, randomized phase I study. *Clin. Vaccine Immunol.* **2016,** *23* (12), 908-917.

111. Kubler-Kielb, J.; Vinogradov, E.; Chu, C.; Schneerson, R., O-Acetylation in the O-specific polysaccharide isolated from Shigella flexneri serotype 2a. *Carbohydr. Res.* **2007,** *342* (3-4), 643-647.

112. Barry, E.; Cassels, F.; Riddle, M.; Walker, R.; Wierzba, T., Vaccines Against Shigella and Enterotoxigenic Escherichia coli: A summary of the 2018 VASE Conference. *Vaccine* **2019,** *37* (34), 4768-4774.

113. Barel, L.-A.; Mulard, L. A., Classical and novel strategies to develop a Shigella glycoconjugate vaccine: from concept to efficacy in human. *Hum. Vaccin. Immunother.* **2019,** *15* (6), 1338-1356.

114. Gauthier, C.; Chassagne, P.; Theillet, F. X.; Guerreiro, C.; Thouron, F.; Nato, F.; Delepierre, M.; Sansonetti, P. J.; Phalipon, A.; Mulard, L. A., Non-stoichiometric O-acetylation of Shigella flexneri 2a O-specific polysaccharide: synthesis and antigenicity. *Org. Biomol. Chem.* **2014,** *12* (24), 4218-4232.

115. Farzam, N.; Ramon-Saraf, R.; Banet-Levi, Y.; Lerner-Geva, L.; Ashkenazi, S.; Kubler-Kielb, J.; Vinogradov, E.; Robbins, J. B.; Schneerson, R., Vaccination with Shigella flexneri 2a

conjugate induces type 2a and cross-reactive type 6 antibodies in humans but not in mice. *Vaccine* **2017,** *35* (37), 4990-4996.

116. Perepelov, A. V.; L'vov, V. L.; Liu, B.; Sof'ya, N. S.; Shekht, M. E.; Shashkov, A. S.; Feng, L.; Aparin, P. G.; Wang, L.; Knirel, Y. A., A similarity in the O-acetylation pattern of the O-antigens of Shigellaflexneri types 1a, 1b, and 2a. *Carbohydr. Res.* **2009,** *344* (5), 687-692.

117. Gaweda, K.; Plazinski, W., Tautomeric and epimeric equilibria of aldo-and ketohexoses studied by the MD simulations and QM calculations. *Carbohydr. Res.* **2019,** *474*, 8-15.

118. Panczyk, K.; Plazinski, W., Pyranose ring puckering in aldopentoses, ketohexoses and deoxyaldohexoses. A molecular dynamics study. *Carbohydr. Res.* **2018,** *455*, 62-70.

119. Del Vigo, E. A.; Marino, C.; Stortz, C. A., Exhaustive rotamer search of the 4C1 conformation of α-and β-d-galactopyranose. *Carbohydr. Res.* **2017,** *448*, 136-147.

120. Kuttel, M. M.; Ravenscroft, N., Conformation and cross-protection in Group B Streptococcus serotype III and Streptococcus pneumoniae serotype 14: a molecular modeling study. *Pharmaceuticals* **2019,** *12* (1), 28.

121. Yang, M.; Simon, R.; MacKerell Jr, A. D., Conformational preference of serogroup B Salmonella O polysaccharide in presence and absence of the monoclonal antibody Se155–4. *J. Phys. Chem. B* **2017,** *121* (15), 3412-3423.

122. Kang, Y.; Barbirz, S.; Lipowsky, R.; Santer, M., Conformational diversity of O-antigen polysaccharides of the Gram-negative bacterium Shigella flexneri serotype Y. *J. Phys. Chem. B* **2014,** *118* (9), 2523-2534.

123. Theillet, F.-X.; Simenel, C.; Guerreiro, C.; Phalipon, A.; Mulard, L. A.; Delepierre, M., Effects of backbone substitutions on the conformational behavior of Shigella flexneri O-antigens: implications for vaccine strategy. *Glycobiology* **2010,** *21* (1), 109-121.

124. Kuttel, M.; Ravenscroft, N.; Foschiatti, M.; Cescutti, P.; Rizzo, R., Conformational properties of two exopolysaccharides produced by Inquilinus limosus, a cystic fibrosis lung pathogen. *Carbohydr. Res.* **2012,** *350*, 40-48.

125. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994,** *101* (5), 4177-4189.

126. Flyvbjerg, H.; Petersen, H. G., Error estimates on averages of correlated data. *J. Chem. Phys.* **1989,** *91* (1), 461-466.

127. Clément, M.-J.; Imberty, A.; Phalipon, A.; Pérez, S.; Simenel, C.; Mulard, L. A.; Delepierre, M., Conformational studies of the O-specific polysaccharide of Shigella flexneri 5a and of four related synthetic pentasaccharide fragments using NMR and molecular modeling. *J. Biol. Chem.* **2003,** *278* (48), 47928-47936.

128. Vyas, N. K.; Vyas, M. N.; Chervenak, M. C.; Johnson, M. A.; Pinto, B. M.; Bundle, D. R.; Quiocho, F. A., Molecular recognition of oligosaccharide epitopes by a monoclonal Fab specific for Shigella flexneri Y lipopolysaccharide: X-ray structures and thermodynamics. *Biochemistry* **2002,** *41* (46), 13575-13586.

129. Vulliez-Le Normand, B.; Saul, F.; Phalipon, A.; Belot, F.; Guerreiro, C.; Mulard, L. A.; Bentley, G., Structures of synthetic O-antigen fragments from serotype 2a Shigella flexneri in complex with a protective monoclonal antibody. *PNAS* **2008,** *105* (29), 9976-9981.

130. Kotloff, K. L.; Platts-Mills, J. A.; Nasrin, D.; Roose, A.; Blackwelder, W. C.; Levine, M. M., Global burden of diarrheal diseases among children in developing countries: Incidence, etiology, and insights from new molecular diagnostic techniques. *Vaccine* **2017,** *35* (49), 6783-6789.

131. Troeger, C.; Blacker, B. F.; Khalil, I. A.; Rao, P. C.; Cao, S.; Zimsen, S. R.; Albertson, S. B.; Stanaway, J. D.; Deshpande, A.; Abebe, Z., Estimates of the global, regional, and national morbidity, mortality, and aetiologies of diarrhoea in 195 countries: a systematic analysis for the Global Burden of Disease Study 2016. *Lancet Infect. Dis.* **2018,** *18* (11), 1211-1228.

132. Baker, S.; The, H. C., Recent insights into Shigella: a major contributor to the global diarrhoeal disease burden. *Curr. Opin. Infect. Dis.* **2018,** *31* (5), 449.

133. Ranjbar, R.; Farahani, A., Shigella: Antibiotic-Resistance Mechanisms And New Horizons For Treatment. *Infect. Drug Resist.* **2019,** *12*, 3137.

134. Ashkenazi, S.; Levy, I.; Kazaronovski, V.; Samra, Z., Growing antimicrobial resistance of Shigella isolates. *J. Antimicrob. Chemother.* **2003,** *51* (2), 427-429.

135. Gaudreau, C.; Pilon, P. A.; Cornut, G.; Marchand-Senecal, X.; Bekal, S., Shigella flexneri with ciprofloxacin resistance and reduced azithromycin susceptibility, Canada, 2015. *Emerg. Infect. Dis.* **2016,** *22* (11).

136. W.H.O., Guidelines for the control of shigellosis, including epidemics due to Shigella dysenteriae type 1. **2005**.

137. Walker, R. I., An assessment of enterotoxigenic Escherichia coli and Shigella vaccine candidates for infants and children. *Vaccine* **2015,** *33* (8), 954-965.

138. DeLaine, B. C.; Wu, T.; Grassel, C. L.; Shimanovich, A.; Pasetti, M. F.; Levine, M. M.; Barry, E. M., Characterization of a multicomponent live, attenuated Shigella flexneri vaccine. *FEMS Pathog. Dis.* **2016,** *74* (5), ftw034.

139. Knirel, Y.; Sun, Q.; Senchenkova, S.; Perepelov, A.; Shashkov, A.; Xu, J., O-Antigen modifications providing antigenic diversity of Shigella flexneri and underlying genetic mechanisms. *Biochemistry-Moscow* **2015,** *80* (7), 901-914.

140. Lindberg, A. A.; Karnell, A.; Weintraub, A., The lipopolysaccharide of Shigella bacteria as a virulence factor. *Rev. Infect. Dis.* **1991,** *13* (Supplement_4), S279-S284.

141. Jennison, A. V.; Verma, N. K., Shigella flexneri infection: pathogenesis and vaccine development. *FEMS Microbiol. Rev.* **2004,** *28* (1), 43-58.

142. Ewing, W. H.; Carpenter, K. P., Recommended designations for the subserotypes of Shigella flexneri. *Int. J. Syst. Evol. Micr.* **1966,** *16* (2), 145-150.

143. Bock, K.; Josephson, S.; Bundle, D. R., Lipopolysaccharide solution conformation: antigen shape inferred from high resolution 1 H and 13 C nuclear magnetic resonance spectroscopy and hard-sphere calculations. *J. Chem. Soc. Perkin Trans. 2* **1982,** (1), 59-70.

144. Kang, Y.; Barbirz, S.; Lipowsky, R.; Santer, M., Conformational diversity of O-antigen polysaccharides of the Gram-negative bacterium Shigella flexneri serotype Y. *J. Phys. Chem. B* **2014,** *118* (9), 2523-2534.

145. Hlozek, J.; Ravenscroft, N.; Kuttel, M. M., The Effects of Glucosylation and O-Acetylation on the Conformation of Shigella Flexneri Serogroup 2 O-Antigen Vaccine Targets. *J. Phys. Chem. B* **2020**.

146. Theillet, F.-X.; Simenel, C.; Guerreiro, C.; Phalipon, A.; Mulard, L. A.; Delepierre, M., Effects of backbone substitutions on the conformational behavior of Shigella flexneri O-antigens: implications for vaccine strategy. *Glycobiology* **2011,** *21* (1), 109-121.

147. Dhara, D.; Kar, R. K.; Bhunia, A.; Misra, A. K., Convergent Synthesis and Conformational Analysis of the Hexasaccharide

Repeating Unit of the O-Antigen of Shigella flexneri Serotype 1d. *Eur. J. Org. Chem.* **2014**, 4577–4584.

148. Von Seidlein, L.; Kim, D. R.; Ali, M.; Lee, H.; Wang, X.; Thiem, V. D.; Canh, D. G.; Chaicumpa, W.; Agtini, M. D.; Hossain, A., A multicentre study of Shigella diarrhoea in six Asian countries: disease burden, clinical manifestations, and microbiology. *PLOS Med.* **2006,** *3* (9).

149. Chompook, P.; Samosornsuk, S.; Von Seidlein, L.; Jitsanguansuk, S.; Sirima, N.; Sudjai, S.; Mangjit, P.; Kim, D. R.; Wheeler, J. G.; Todd, J., Estimating the burden of shigellosis in Thailand: 36-month population-based surveillance study. *Bulletin of the W.H.O.* **2005,** *83*, 739-746.

150. Boutet, J.; Blasco, P.; Guerreiro, C.; Thouron, F.; Dartevelle, S.; Nato, F.; Cañada, F. J.; Ardá, A.; Phalipon, A.; Jiménez-Barbero, J., Detailed Investigation of the Immunodominant Role of O-Antigen Stoichiometric O-Acetylation as Revealed by Chemical Synthesis, Immunochemistry, Solution Conformation and STD-NMR Spectroscopy for Shigella flexneri 3a. *Chem. Eur. J.* **2016,** *22* (31), 10892-10911.

151. Carlin, N.; Wehler, T.; Lindberg, A., Shigella flexneri O-antigen epitopes: chemical and immunochemical analyses reveal that epitopes of type III and group 6 antigens are identical. *Infect. Immun.* **1986,** *53* (1), 110-115.

152. Wang, J.; Lan, R.; Knirel, Y. A.; Luo, X.; Sof'ya, N. S.; Shashkov, A. S.; Xu, J.; Sun, Q., Serological identification and prevalence of a novel O-antigen epitope linked to 3-and 4-O-acetylated rhamnose III of lipopolysaccharide in Shigella flexneri. *J. Clin. Microbiol.* **2014,** *52* (6), 2033-2038.

153. Onodera, N. T.; Ryu, J.; Durbic, T.; Nislow, C.; Archibald, J. M.; Rohde, J. R., Genome sequence of Shigella flexneri serotype 5a strain M90T Sm. *J Bacteriol Res* **2012**.

154. Kuttel, M. M.; Timol, Z.; Ravenscroft, N., Cross-protection in Neisseria meningitidis serogroups Y and W polysaccharides: A comparative conformational analysis. *Carbohydr. Res.* **2017,** *446*, 40-47.

155. Phillips, J. C.; Braun, R.; Wang, W.; Gumbart, J.; Tajkhorshid, E.; Villa, E.; Chipot, C.; Skeel, R. D.; Kale, L.; Schulten, K., Scalable molecular dynamics with NAMD. *J. Comput. Chem.* **2005,** *26* (16), 1781-1802.

156. Stone, J. E.; Phillips, J. C.; Freddolino, P. L.; Hardy, D. J.; Trabuco, L. G.; Schulten, K., Accelerating molecular modeling applications with graphics processors. *J. Comput. Chem.* **2007,** *28* (16), 2618-2640.

157. Jorgensen, W. L.; Chandrasekhar, J.; Madura, J. D.; Impey, R. W.; Klein, M. L., Comparison of simple potential functions for simulating liquid water. *J. Chem. Phys.* **1983,** *79* (2), 926-935.

158. Kuttel, M. M.; Ståhle, J.; Widmalm, G., CarbBuilder: Software for building molecular models of complex oligo-and polysaccharide structures. *J. Comput. Chem.* **2016,** *37* (22), 2098-2105.

159. Humphrey, W.; Dalke, A.; Schulten, K., VMD: visual molecular dynamics. *J. Mol. Graphics.* **1996,** *14* (1), 33-38.

160. Allen, M. P.; Tildesley, D. J., *Computer simulation of liquids*. Oxford university press: 2017.

161. Martyna, G. J.; Tobias, D. J.; Klein, M. L., Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* **1994,** *101* (5), 4177-4189.

162. Cross, S.; Kuttel, M. M.; Stone, J. E.; Gain, J. E., Visualisation of cyclic and multi-branched molecules with VMD. *J. Mol. Graphics Modell.* **2009,** *28* (2), 131-139.

163. Gimeno, A.; Valverde, P.; Ardá, A.; Jiménez-Barbero, J., Glycan structures and their interactions with proteins. A NMR view. *Current Opinion in Structural Biology* **2020,** *62*, 22-30.

164. Dharmasena, M. N.; Osorio, M.; Takeda, K.; Stibitz, S.; Kopecko, D. J., Stable chromosomal expression of Shigella flexneri 2a and 3a O-antigens in the live Salmonella oral vaccine vector Ty21a. *Clin. Vaccine Immunol.* **2017,** *24* (12), e00181-17.

165. Lin, J.; Smith, M. A.; Benjamin, W. H.; Kaminski, R. W.; Wenzel, H.; Nahm, M. H., Monoclonal antibodies to Shigella lipopolysaccharide are useful for vaccine production. *Clin. Vaccine Immunol.* **2016,** *23* (8), 681-688.

166. Nahm, M. H.; Yu, J.; Weerts, H. P.; Wenzel, H.; Tamilselvi, C. S.; Chandrasekaran, L.; Pasetti, M. F.; Mani, S.; Kaminski, R. W., Development, interlaboratory evaluations, and application of a simple, high-throughput Shigella serum bactericidal assay. *mSphere* **2018,** *3* (3), e00146-18.

167. Lipsitz, R. S.; Tjandra, N., Residual dipolar couplings in NMR structure analysis. *Annu. Rev. Biophys. Biomol. Struct.* **2004,** *33*, 387-413.